Organisation, variabilité et contrôle épigénétique des gènes d'ARNr 5S chez Arabidopsis thaliana

Lauriane Simon

CLERMONT UNIVERSITE **Année 2016**

N° D.U : 2783

École doctorale

Sciences de la Vie, Santé, Agronomie, Environnement

N° d'ordre : 710

Thèse

Présentée à l'Université Blaise Pascal pour l'obtention du grade de

DOCTEUR D'UNIVERSITÉ

Spécialité : Physiologie et Génétique Moléculaires

Soutenue le 19 décembre 2016

Lauriane SIMON

Organization, variability and epigenetic control of 5S rRNA genes in *Arabidopsis thaliana*

Présidente :	Chantal VAURY, GReD, Clermont-Ferrand
Rapporteurs :	Julio SAEZ-VASQUEZ, LGDP, Perpignan
	Moussa BENHAMED, IPS2, Paris-Saclay
Directrice de thèse :	Aline PROBST, GReD, Clermont-Ferrand

Génétique, Reproduction & Développement
UMR CNRS 6293 - Clermont Université - INSERM U1103
10 avenue Blaise Pascal 63 178 Aubière

CLERMONT UNIVERSITE **Année 2016**

N° D.U : 2783

École doctorale

Sciences de la Vie, Santé, Agronomie, Environnement

N° d'ordre : 710

Thèse

Présentée à l'Université Blaise Pascal pour l'obtention du grade de

DOCTEUR D'UNIVERSITÉ

Spécialité : Physiologie et Génétique Moléculaires

Soutenue le 19 décembre 2016

Lauriane SIMON

Organisation, variabilité et contrôle épigénétique des gènes d'ARNr 5S chez *Arabidopsis thaliana*

Présidente :	Chantal VAURY, GReD, Clermont-Ferrand
Rapporteurs :	Julio SAEZ-VASQUEZ, LGDP, Perpignan
	Moussa BENHAMED, IPS2, Paris-Saclay
Directrice de thèse :	Aline PROBST, GReD, Clermont-Ferrand

Génétique, Reproduction & Développement
UMR CNRS 6293 - Clermont Université - INSERM U1103
10 avenue Blaise Pascal 63 178 Aubière

Voici venu le temps des remerciements, dernière étape avant de rendre le manuscrit et d'officiellement finir cette thèse.

Ma première pensée va à ma chef Aline, sans qui cette thèse n'aurait pas été possible. Merci de m'avoir donné ma chance en M2 et d'avoir obtenu une bourse de thèse de la région pour moi. Merci pour ta gentillesse, ton investissement et ton soutien, notamment au moment où la motivation me faisait défaut. Merci aussi pour ton expertise et ton esprit critique, j'espère un jour être une chef comme toi.

Merci aussi à Christophe, mon deuxième chef, merci de m'avoir parlé lors des portes ouvertes de l'université et de m'avoir donné l'envie de venir faire mon Master à Clermont. Merci d'avoir été présent lors des discussions sur le 5S et d'avoir été le troisième avis qui nous a souvent aidées à trancher. Merci pour tes conseils et pour tes « relations » qui m'ont permis d'avancer sur le projet.

Merci donc à Hervé Duborjal, Yannick Bidet, et la plateforme de séquençage Gentyane pour leurs aides dans la préparation des expériences de séquençages à haut débit.

Merci aussi à tous les membres de mon équipe, présents et passés et notamment :

Merci à Sam pour sa grande gentillesse, pour « salade de fruits » fredonnée dans les couloirs et France inter. Désolée de t'avoir toujours appelé lorsque je me retrouvais face à des objets cassés ou en panne.

Merci à Sylvie et Sylvianne, pour les repas du midi, les pauses à quatre heures et particulièrement à Sylvie pour les « éclaircissements » sur le 5S.

Merci à Manu, pour les discussions parfois animées, le partage de nos goûts en séries. Merci d'avoir été le Zizanix de l'équipe et d'être toujours honnête.

Merci à Céline, pour le partage aussi bien au labo qu'en dehors, merci d'avoir été là, de m'avoir appris un peu de ta rigueur et fait partager ton expérience. J'espère un jour être aussi ordonnée et organisée que toi mais j'ai encore un long chemin à faire. J'espère que tu trouveras bientôt un peu de stabilité, tu le mérites.

Merci enfin à Maxime dit « Max » jeune padawan. Merci d'être arrivé au moment où je me retrouvais un peu seule dans l'équipe, merci pour les discussions sur les jeux vidéo, les lasers games, le pain délicieux, les quiches au thon et pour ces moments d'écoute, de partage et de ras le bol. Courage pour ta fin de thèse.

Merci aussi à Axel, Mélanie et Matthias, pour m'avoir accueillie dans l'équipe et m'avoir fait une place. Merci pour tous les moments de rigolades, de folies et de partages, merci pour l'aide et le soutien dans les moments difficiles. Le labo m'a semblé vide quand vous êtes partis.

Je voudrais aussi remercier les gens du couloir plantes, pour m'avoir supportée pendant ces 3 ans et avoir fait en sorte que ces 3 années resteront un très bon souvenir pour moi.

Merci à Margaux et à Pierre, pour tous les moments partagés, les pauses Nutella, et les messes basses dans le bureau.

Merci aussi aux autres amis thésards, Manue, Claire et Angeline, merci d'avoir été là, merci pour les restos, les apéros, les courts métrages et les cinés et bien sûr pour ces 14 février partagés. Merci pour tous ces moments de détente même si on aurait dû en faire plus souvent. Merci d'avoir partagé ces 3 ans avec moi.

Merci enfin à ma famille, merci à mes parents et ma sœur, merci d'avoir été là, de m'avoir soutenue et supportée, et de vous être intéressés à ce que je faisais. Merci pour votre amour.

Abstract

5S rRNA is an integral component of the ribosome and is essential for protein synthesis. 5S rRNA expression is therefore tightly regulated to suit the cellular requirements. 5S rRNA is encoded by gene arrays organized in tandem repeats mainly situated in the pericentromeric regions of chromosomes 3, 4 and 5 in the *Arabidopsis thaliana* Col-0 genome. Full genome assembly remains challenging in these highly repeated regions and impedes further investigation of their organization, dynamics and epigenetic regulation.

In this thesis, we provide information on 5S rRNA genes in *Arabidopsis thaliana*. We confirmed specific DNA signatures by deep sequencing and define chromosome-specific polymorphisms and new reference sequences. We also studied the epigenetic status of the 5S rRNA genes showing that these genes are enriched in repressive marks whereas the *loci* on chromosome 4 and 5 also display peculiar histone modifications and variants characteristic of transcriptional activity. We report that 5S rDNA *loci* are highly dynamic within the Arabidopsis species with high level of variation in global copy number and cluster proportion between ecotypes. Using the L*er* ecotype, in which reorganization events were recorded, and specific mutant backgrounds, we identified chromosome 5 rDNA *locus* as a major source of variation and observed altered chromatin organization in nuclear space as a consequence of 5S rDNA *locus* variation. Finally, we suggest that differences in chromatin states at the two main 5S rDNA *loci* can lead to differential usage of 5S rRNA genes in different ecotypes.

The analysis provides evidence for a role of chromatin organization in transcriptional regulation and 5S rDNA organization.

Résumé

L'ARNr 5S est une partie intégrante du ribosome indispensable pour la synthèse des protéines. Cette fonction essentielle nécessite une régulation fine de l'expression des ARNr 5S en fonction des exigences cellulaires. L'ARNr 5S est codé par des gènes organisés en répétitions en tandem principalement situés dans les régions péricentromériques des chromosomes 3, 4 et 5 dans le génome Col-0 d'*Arabidopsis thaliana*. L'étude approfondie de l'organisation, la dynamique et la régulation épigénétique des gènes d'ARNr 5S reste cependant difficile du fait de l'absence d'assemblage de ces régions hautement répétées dans la séquence du génome d'Arabidopsis disponible dans TAIR10.

Dans cette thèse, nous apportons de nouvelles informations sur les gènes d'ARNr 5S chez *Arabidopsis*. Nous avons confirmé les signatures d'ADN spécifiques par séquençage haut débit, identifié des polymorphismes spécifiques des régions chromosomiques portant des ADNr 5S et ainsi défini de nouvelles séquences de référence. Nous avons également étudié le statut épigénétique des gènes ARNr 5S, montrant que ces gènes sont globalement enrichis en marques répressives. Les *loci* situés sur les chromosomes 4 et 5 présentent également des modifications post-traductionnelles d'histones et des variants d'histone particuliers et caractéristiques d'une activité de transcription. Nous avons aussi montré que les *loci* d'ADNr 5S sont très dynamiques au sein de l'espèce *Arabidopsis* avec des variations entre différent écotypes du nombre de copies de gènes d'ARNr 5S et de l'importance relative de chaque cluster. En utilisant l'écotype L*er* et des mutants spécifiques, nous montrons que le *locus* du chromosome 5 est une source majeure de réarrangements chromosomiques qui affectent l'organisation de la chromatine dans le noyau. Enfin, nous suggérons que des variations de l'état chromatinien des gènes d'ARNr 5S peuvent conduire à une différence de transcription dans les différents écotypes.

Cette étude permet d'établir un rôle de l'organisation de la chromatine dans la régulation transcriptionnelle et l'organisation des gènes d'ARNr 5S.

Sommaire

Liste des Abréviations

3C : Chromosome Conformation Capture

5-AZA : 5-aza-2′-deoxycytosine

AGO4 : ARGONAUTE 4

ARN : Acide RiboNucléique

ARNm : ARN messager

ARNnc : ARN non codant

ARNr : ARN ribosomique

ARNsn : small nuclear ARN

ARNt : ARN de transfert

ATM : Ataxia-Telangiectasia Mutated

ATRX : Alpha Thalassemia-mental Retardation X-linked

ATX : ARABIDOPSIS HOMOLOGY OF TRITHORAX

ATXR : ARABIDOPSIS TRITHORAX-RELATED

BAC : Bacterial Artificial Chromosome

bwa : Burrows-Wheeler Aligner

CABIN1 : CAlcineurin BINding protein 1

CAF-1 : Chromatin assembly factor 1

ChIP : Chromatin Immuno-Precipitation

CHR20 : Chromatin Remodelling 20

CMT3 : CHROMOMETHYLASE 3

Col-0 : Columbia

CS : Chromatin State

Cvi : Cape Verde Island

DAPI : 4′,6′-diamidino-2-phenylindole

DAXX : Death-domain Associated protein

DCL3 : DICER-Like 3

DDM1 : DEFICIENT IN DNA METHYLATION 1

DNA : Deoxyribonucleic Acid

DRM2 : DOMAINS REARRANGED METHYLTRANSFERASE 2

EdU : 5-Ethynyl-2’-deoxyUridine

EI : Exon-inclusion

ES : Exon-skipping

ESPript : Easy Sequencing in PostScript

ETS : External Transcribed Spacer

FACT : Facilitates Chromatin Transcription

FAS : FASciata

FISH : Fluorescence *In Situ* Hybridization

FLASH : Fast Length Adjustment of Short Reads

Fob1 : FOrk Blocking protein 1

GFP : Green Fluorescent Protein

HDA6 : HISTONE DEACETYLASE 6

HIR : HISTONE REGULATOR

HIRA : HIstone Regulator A

HTR : HISTONE THREE RELATED

HU : Hydroxyurea

ICR : Internal Control Region

IGS : Intergenic spacer

ITS : Internal Transcribed Spacer

Kas : Kashmir1

KEE : KNOT engaged element

La-0 : Landsberg

L*er* : Landsberg *erecta*

LNA : Locked nucleic acids

m5C : 5-methylcytosine

MET1 : METHYLTRANSFERASE 1

MS : Murashige & Skoog

MSI1 : MULTICOPY SUPPRESSOR OF IRA1

NASC : Nottingham Arabidopsis Stock Center

NOR : Nucleolus Organizer Regions

OCS : Octopine synthase

Pol : ARN polymérase

PTM : Post-Translational Marks

RdDM : RNA-directed DNA methylation

RDR2 : RNA-DEPENDENT RNA POLYMERASE 2

RNP : Ribonucleoprotein particle

ROS1 : REPRESSOR OF SILENCING 1

RT-qPCR : Quantitative RT-PCR Reverse Transcription-quantitative Polymerase Chain Reaction

SDG2 : SET DOMAIN GROUP 2

SINE : Short Interspersed Nuclear Element

SL1 : Selectivity factor1

SSA : Single-strand annealing

SUVH : SUPPRESSOR OF VARIEGATION (SU(VAR)3-9) HOMOLOGS

TBP : TATA-Binding Protein

TFIIIA : Transcription factor III A

TFIIIB : Transcription factor III B

TFIIIC : Transcription factor III C

TSI : Transcriptional Silent Information

TTS : Transcription Termination Site

UBF : Upstream Binding Factor

UBN1/2 : UBiNuclein 1/2

VIM1 : VARIATION IN METHYLATION 1

WS : Wassileskija

YAC : Yeast Artificial Chromosome

Liste des illustrations

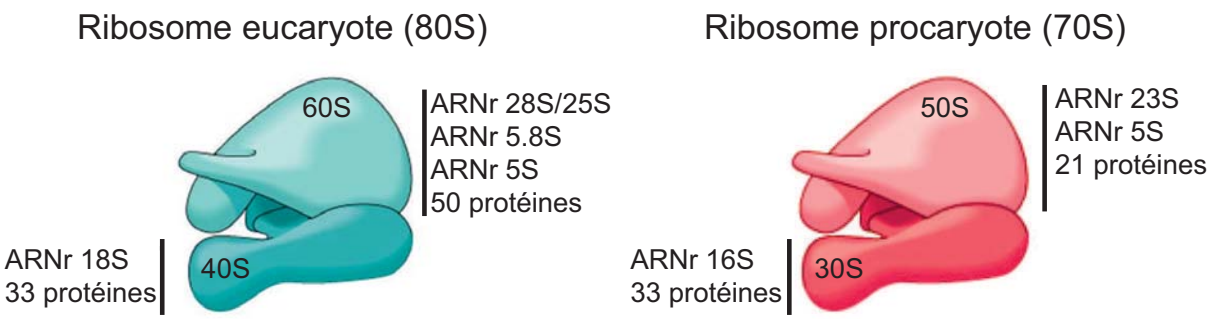


Figure 1 : Composition des ribosomes eucaryotes et procaryotes.
Chez les plantes, l'ARNr 28S est remplacé par l'ARNr 25S.

Promoteurs ARN Pol III	gènes	Motifs
Type I	ARNr 5S	Promoteur interne : Boites A, C Elément Intermédiaire (EI)
Type II	ARNt	Promoteur interne : Boites A, C
Type III	7SK, U6, …	Eléments ESP, ESD, Boîte TATA

Figure 2 : Les trois types de promoteur reconnus par l'ARN polymérase III.
Les séquences promotrices du promoteur de type I incluent les boîtes A et C ainsi qu'un élément intermédiaire. Dans les promoteurs de type II, la boîte C est remplacée par une boîte B. Enfin, les promoteurs de type III sont composés des éléments de séquence distal (ESD), proximal (ESP) et d'une boîte TATA.

Synthèse bibliographique

I. Le ribosome

La synthèse protéique, la traduction, est un processus universel et central de la machinerie cellulaire puisqu'elle permet la traduction des protéines. L'acteur principal de la traduction est le ribosome, il forme la chaîne polypeptidique en mettant en relation l'ARN messager (ARNm) à traduire, les ARN de transfert (ARNt) portant les acides aminés et divers facteurs protéiques. De par sa fonction essentielle dans la vie cellulaire, le ribosome est retrouvé chez tous les organismes, aussi bien chez les procaryotes que chez les eucaryotes.

Le ribosome est un complexe ribonucléoprotéique composé de protéines et d'ARN ribosomiques (ARNr) (qui représentent environ 80% des ARNs totaux de la cellule). Le ribosome est toujours composé d'une grande sous unité qui porte l'activité catalytique et d'une petite sous unité qui décode l'information portée par l'ARNm.

Bien que le ribosome soit présent chez tous les organismes vivants, sa composition en protéines et en ARNr est variable (Figure 1). Chez les procaryotes, le ribosome (70S) est composé d'une grande sous unité (50S) comprenant ~34 protéines, les ARNr 5S et 23S et d'une petite sous unité (30S) comprenant ~21 protéines et l'ARNr 16S. Chez les eucaryotes, le ribosome (80S) est composé de la grande sous unité (60S) qui se compose de ~49 protéines et de 3 molécules d'ARNr : l'ARNr 5S, le 5.8S et le 28S (25S chez les plantes). La petite sous unité (40S) se compose de ~33 protéines et de l'ARNr 18S (John L. Woolford and Warner, 1991; Raué and Planta, 1991). Tous ces composants sont associés en particules pré-ribosomales au niveau du nucléole qui est une structure nucléaire où sont synthétisés et maturés les ARNr 5.8S, 25S et 18S (Shaw and Jordan, 1995). Les particules pré-ribosomales sont ensuite exportées dans le cytoplasme où elles forment un ribosome mature.

La formation du ribosome fait intervenir trois ARN polymérases : l'ARN polymérase I (Pol I) pour la transcription des gènes d'ARNr 18S, 5.8S et 23S, l'ARN polymérase II (Pol II) pour les protéines ribosomales, et l'ARN polymérase III (Pol III) qui transcrit l'ARNr 5S.

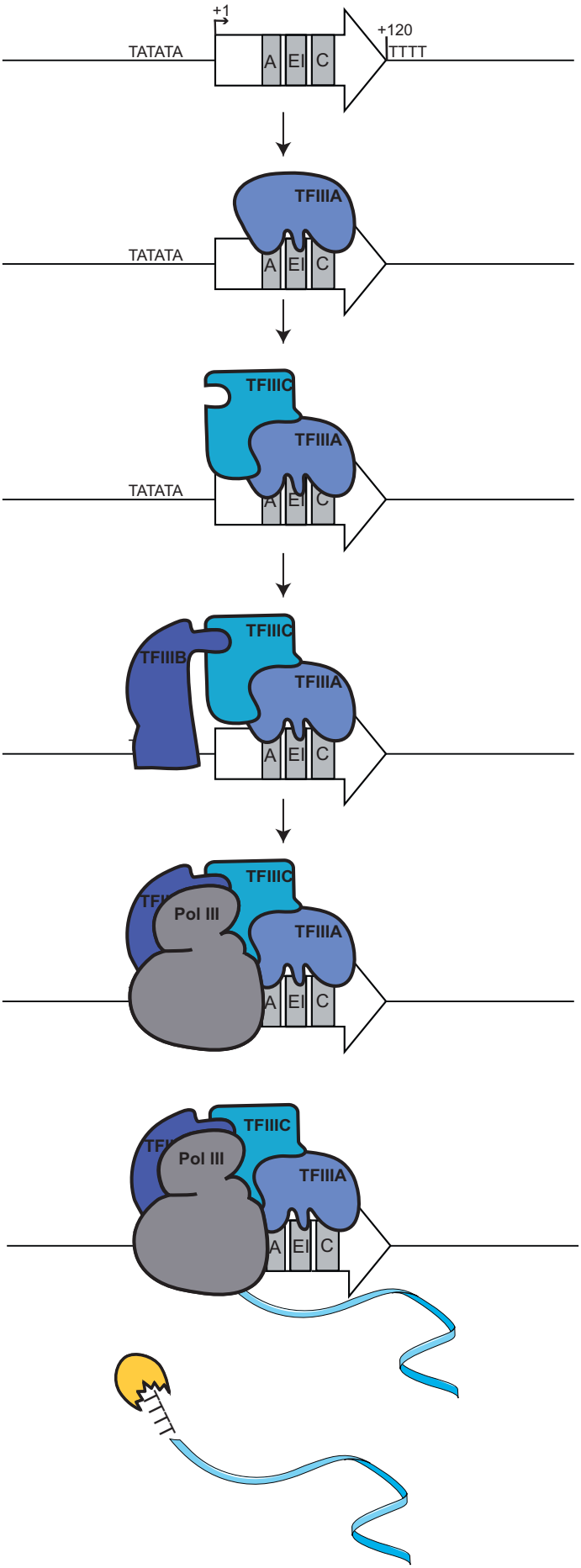


Figure 3 : Assemblage du complexe de transcription sur un gène d'ARNr 5S. La fixation du facteur de transcription TFIIIA sur le promoteur interne permet la liaison de TFIIIC. La présence de TFIIIA et TFIIIC permet de recruter TFIIIB. Enfin, TFIIIB permet le recrutement de l'ARN polymérase III (d'après la thèse de Julien Douet).

Chez *Arabidopsis thaliana* (*A. thaliana*), chacune des 80 protéines ribosomales est codée par des gènes distincts. De deux à sept gènes paralogues codent pour chaque protéine et ces paralogues montrent entre 65% et 100% d'identité au niveau protéique (Barakat et al., 2001). L'analyse de la composition protéique des ribosomes a démontré qu'une vaste majorité de ces protéines paralogues sont incorporées dans les ribosomes (Hummel et al., 2012) et que la composition du ribosome en ces protéines peut varier avec le stade de développement, le tissu ou encore en fonction de *stimuli* environnementaux (Byrne, 2009; Hummel et al., 2012; Szick-Miranda and Bailey-Serres, 2001). Les mécanismes de régulations de l'expression des gènes codant pour les protéines paralogues et l'étude de leur fonction dans le ribosome restent à découvrir.

De par sa constitution ribonucléoprotéique et son rôle central dans la vie cellulaire, la formation du ribosome doit être finement régulée.

II. Les ARNr

A. Mécanisme de transcription et de maturation

Chez les eucaryotes, l'ARN polymérase III assure la transcription de l'ARNr 5S, des ARNt, des ARNsn (small nuclear RNA), de certains ARNnc (ARN non codants) ou encore des SINEs (Short Interspersed Nuclear Element) (Dieci et al., 2007; Paule and White, 2000).

L'ARN polymérase III reconnaît et se fixe sur trois types de promoteurs (Figure 2): le type 1 présent dans les ARNr 5S, le type 2 caractéristique des ARNt et le type 3 caractéristique des ARNsn ou des SINEs (Paule and White, 2000).

La transcription par l'ARN polymérase III nécessite toujours l'intervention de deux facteurs de transcription : TFIIIB et TFIIIC. La synthèse des ARNr 5S requiert en plus l'intervention d'un facteur de transcription spécifique appelé TFIIIA (Engelke et al., 1980).

L'activité de l'ARN polymérase III lors de la transcription des ARNr 5S a été bien décrite chez *S. cerevisiae*. Premièrement, le facteur TFIIIA se fixe sur le promoteur interne puis il recrute le facteur TFIIIC qui permet à son tour de recruter le facteur TFIIIB. Un des composants de TFIIIB, le peptide TBP (TATA-Binding Protein), se fixe sur une boîte TATA-like en amont de la séquence transcrite ce qui permet le recrutement de l'ARN polymérase III (Fig

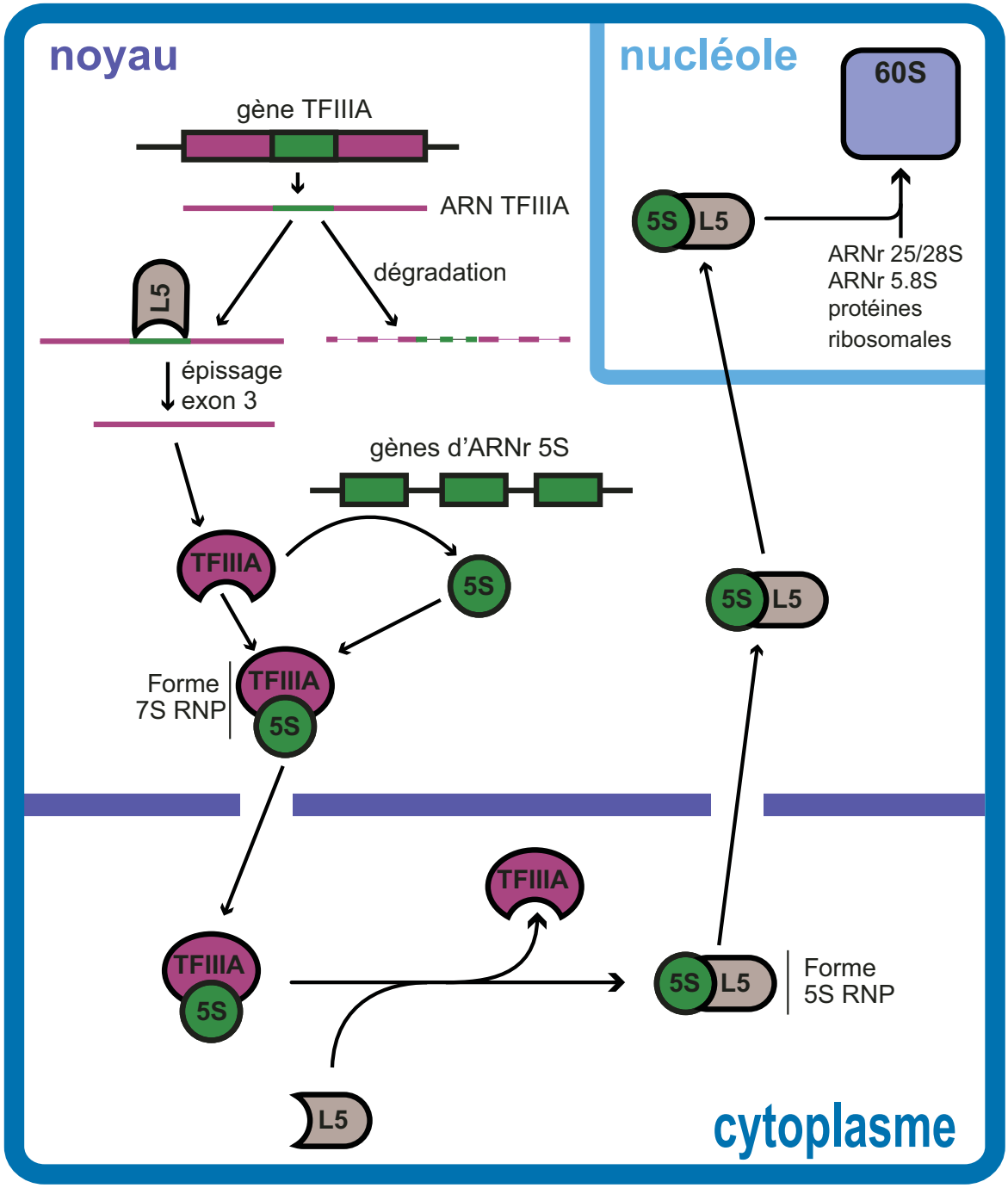


Figure 4 : Mécanismes de synthèse de la protéine TFIIIA, de la transcription et du transport de l'ARNr 5S.

L'exon 3 du gène codant pour TFIIIA correspond à une partie de l'ARNr 5S. Cette séquence 5S est reconnue par la protéine L5 ce qui va favoriser l'excision de cet exon et permettre la formation d'une protéine TFIIIA pleine taille. TFIIIA va ensuite participer à la transcription des ARNr 5S puis TFIIIA lie spécifiquement l'ARNr 5S et assure son export vers le cytoplasme. L'ARNr 5S est ensuite pris en charge par la protéine L5 pour être importé dans le nucléole où il sera associé à la grande sous-unité ribosomale 60S. Si peu de molécules d'ARNr 5S sont présentes dans la cellule, les protéines L5 libres vont se fixer sur l'exon 3 de l'ARN TFIIIA ce qui va favoriser l'épissage et la production d'une protéine TFIIIA et donc la synthèse d'ARNr 5S. A l'inverse si beaucoup de molécules d'ARNr 5S sont présentes dans la cellule, les protéines L5 vont être associées aux ARNr 5S et ne se fixeront pas sur l'ARN TFIIIA, moins de TFIIIA et donc moins d'ARNr 5S seront produit.

ure 3) (Bieker et al., 1985; Kassavetis et al., 1991; Setzer and Brown, 1985).

A la fin de la transcription, l'ARN polymérase III reconnaît une répétition de thymine comme le signal de fin de transcription (Bogenhagen and Brown, 1981). Chez la plante modèle *A. thaliana* la présence de quatre thymines est suffisante et nécessaire *in vitro* pour mettre fin à la transcription (Cloix et al., 2003). L'extrémité 3' de l'ARNr 5S est par la suite maturée par des exonucléases afin d'obtenir un ARNr 5S mature de 120 bases (Ciganda and Williams, 2011).

Les ARNr 18S, 5.8S, 25S sont transcrits par l'ARN polymérase I en un long précurseur polycistronique (pré-ARNr 45S) qui va par la suite être clivé par des exo- et endonucléases pour donner les trois ARNr (revue : Weis et al., 2015).

Chez l'homme, la transcription par l'ARN Pol I fait intervenir trois composants majeurs: l'ARN polymérase I, SL1 (selectivity factor1) et UBF (upstream binding factor). Le complexe SL1 interagit avec le promoteur du gène d'ARNr 45S et stabilise la fixation d'UBF sur le promoteur. SL1 recrute par la suite l'ARN Pol I.

B. Prise en charge des ARNr dans la cellule

Le facteur TFIIIA a une double fonction. D'une part il participe à la transcription de l'ARNr 5S et d'autre part, il prend en charge l'ARN 5S. Cette association forme le complexe 7S RNP qui va permettre l'exportation et le stockage de l'ARNr 5S dans le cytoplasme. Afin de former les particules pré-ribosomales la protéine L5 remplacera le facteur TFIIIA et prendra en charge l'ARNr 5S afin d'assurer son retour vers le noyau et plus précisément dans le nucléole (Figure 4) (Allison et al., 1991; Guddat et al., 1990). La protéine L5 a été identifiée chez *A. thaliana* par la recherche dans les bases de données de protéines comportant les 9 doigts de zinc caractéristiques de cette protéine. La protéine L5 montre une localisation cytoplasmique et nucléaire, mais avec un enrichissement dans le nucléole (Mathieu et al., 2003b). Le facteur TFIIIA ainsi libéré va pouvoir de nouveau participer à la transcription de l'ARNr 5S dans le noyau. La quantité et la disponibilité des protéines L5 et TFIIIA dans la cellule vont donc contribuer à la régulation de l'expression de l'ARNr 5S (Layat et al., 2012; Mathieu et al., 2003b; Pittman et al., 1999).

Chez *A. thaliana*, le gène TFIIIA est lui-même finement régulé par épissage alternatif de l'exon 3 qui génère deux transcrits. Le transcrit avec exon 3 (EI : exon-inclusion) n'est pas traduit et est dégradé dans la cellule. Le transcrit sans exon 3 (ES : exon-skipping) est traduit

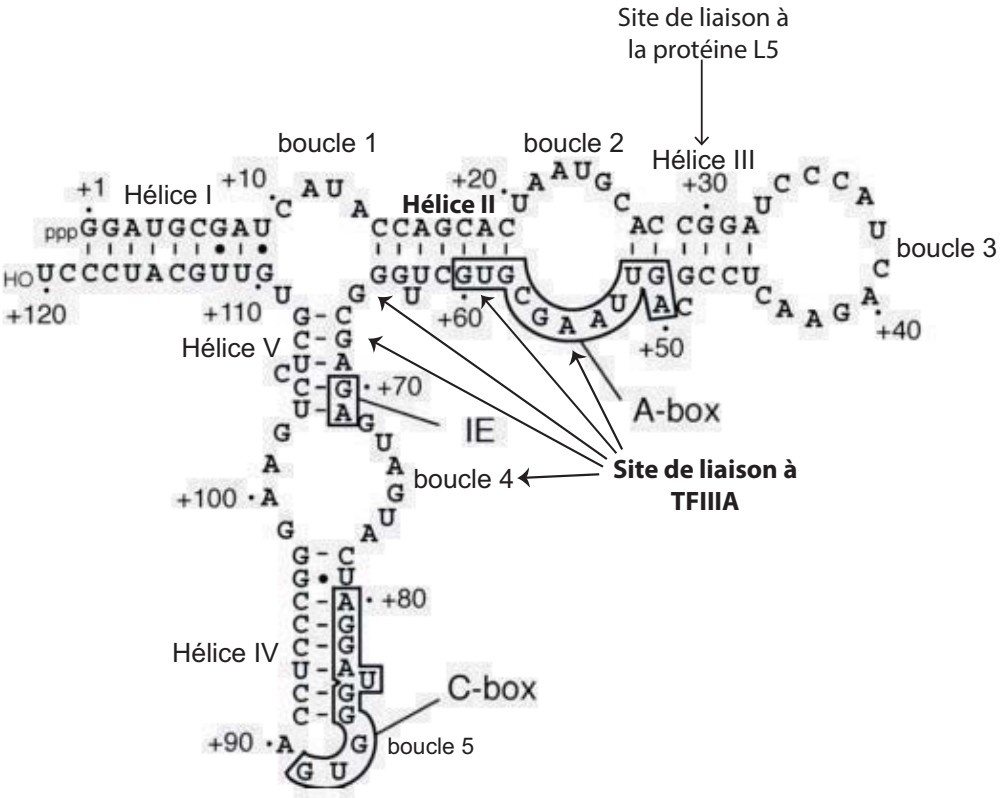


Figure 5 : Structure secondaire de l'ARNr 5S chez *Arabidopsis thaliana*.
Les sites importants pour la liaison des protéines TFIIIA et L5 ainsi que le promoteur interne sont indiqués. (Cloix et al., 2003)

et donne la protéine TFIIIA de pleine taille. L'exon 3 possède une très forte identité de séquence et de structure avec la séquence du gène d'ARNr 5S incluant les séquences nécessaires à la liaison à la protéine L5 (hélice III) (Hammond et al., 2009). Cet exemple d'exonisation d'un gène d'ARNr 5S favoriserait l'épissage de l'exon 3 de l'ARN TFIIIA en cas de liaison avec la protéine L5. Lorsque beaucoup d'ARNr 5S sont présents dans la cellule, les protéines L5 vont se fixer sur ceux-ci et pas à l'exon 3 de TFIIIA et donc seul le transcrit EI est synthétisé. Celui ci est dégradé ce qui diminue le taux de protéine TFIIIA dans la cellule et donc le taux d'ARNr 5S. Si les protéines L5 sont libres dans la cellule car il y a peu d'ARNr 5S présents, elles viennent se lier à l'exon 3 ce qui en favoriserait l'épissage et permettrait la synthèse du transcrit ES et donc la production de protéine TFIIIA pleine taille activant la transcription de l'ARNr 5S (Fu et al., 2009; Hammond et al., 2009).

Les ARNr 18S, 5.8S et 23S sont eux directement transcrits et maturés dans le nucléole où ils sont par la suite intégrés aux particules pré-ribosomales.

Peu de choses sont connues sur la maturation de l'ARNr 45S chez *A. thaliana*. Chez l'homme et la levure, la maturation de l'ARNr 45S commence par la production de l'ARNr 18S par l'activité d'endo- et/ou exonucléases. Ensuite, des nucléases maturent les extrémités des ARNr 5.8S et 25/28S. Ce processus fait intervenir beaucoup de protéines et facteurs d'épissage pour une maturation correcte de l'ARNr 45S (Henras et al., 2015).

C. L'ARNr 5S: structure et fonction

L'ARNr 5S est un petit ARN de 120 nucléotides qui présente une structure secondaire conservée au cours de l'évolution (Figure 5, (Barciszewska et al., 2000)). Il comporte 5 hélices (I à V) et 5 boucles (A à E). Les hélices III, IV, V et les boucles A/B/C servent de liaison au facteur de transcription TFIIIA et l'hélice III à la liaison à la protéine L5. L'hélice I est impliquée dans la stabilisation de l'ARNr 5S et dans son intégration au ribosome (Dinman, 2005).

Bien que l'ARN 5S ait été utilisé comme modèle dans l'étude de la conformation des ARNs, des interactions ARN/protéines ou comme marqueur phylogénétique (Alonso-Blanco et al., 2016; Alves-Costa et al., 2008; Perina et al., 2011; Waminal et al., 2013), sa fonction au sein du ribosome reste encore mal connue. Néanmoins, l'ARNr 5S est pressenti comme étant un acteur central de la transmission de l'information en agissant comme un transducteur physique de l'information entre les différents centres fonctionnels du ribosome

(Smith et al., 2001) et comme stabilisateur de la grande sous unité du ribosome en amenant et stabilisant différentes régions des ARNr dans le centre catalytique du ribosome (Holmberg and Nygård, 2000; Khaitovich and Mankin, 1999).

D. Hétérogénéité des transcrits d'ARNr

Chez *A. thaliana*, le séquençage de quelques centaines de produits RT-PCR a montré qu'il existerait au moins quatre types de transcrits issus des gènes d'ARNr 5S (Cloix et al., 2002; Vaillant et al., 2006) :

- des ARNr 5S de 120 bases qui sont inclus dans le ribosome. Ces ARNs ont été définis comme les ARNr 5S « majoritaires » (Cloix et al. 2002) et leur séquence constitue la **séquence consensus** de référence des ARNr 5S.

- une fraction des ARNr 5S de 120 bases portent 1 à 2 polymorphismes par rapport à la séquence consensus et sont qualifiés d'ARN 5S « minoritaires ». Ces ARN ne sont exprimés qu'à certains stades du développement ou chez certains mutants affectant l'organisation de la chromatine et la régulation épigénétique (Mathieu et al., 2003a; Vaillant et al., 2007). Des analyses de polysomes ont montré que ces ARNr minoritaires portant un à deux polymorphismes sont incorporés au ribosome et semblent donc être fonctionnels (Cloix et al., 2002).

- un transcrit aberrant de 210 bases (5S-210) qui n'est exprimé que dans certains mutants et qui est considéré comme un marqueur de la levée d'une répression transcriptionnelle des gènes d'ARNr 5S (Mathieu et al., 2003a). Ce transcrit pourrait être le résultat d'un 'read-through' transcriptionnel généré par l'ARN Pol III (Blevins et al., 2009).

- un transcrit aberrant de 140 bases (5S-140) présentant une délétion de 70 nucléotides en comparaison à l'ARNr 5S-210. Cet ARN pourrait provenir de gènes d'ARNr 5S tronqués (Vaillant et al., 2006).

Peu de variants pour les ARNr 5.8S et 25S ont été identifiés jusqu'à présent. Par contre, le séquençage Sanger de 34 clones d'ARNr 18S a montré que la majorité des transcrits 18S portaient une mutation par rapport à la séquence consensus et qu'il existait un transcrit plus court qui correspond à la transcription de gènes d'ARNr 18S ayant subi une délétion de 270pb correspondant aux positions 456-725 du transcrit pleine taille (Mentewab et al., 2011).

Les gènes d'ARNr 45S existent sous la forme de 4 à 5 variants correspondant à des variations au niveau de leur séquence 3' (voir chapitre "Hétérogénéité des séquences d'ADNr et d'ARNr") (Havlová et al., 2016; Pontvianne et al., 2010). Les niveaux d'expression de ces variants diffèrent en fonction du tissu étudié et dans certains mutants (Abou-Ellail et al., 2011; Chandrasekhara et al., 2016; Pontvianne et al., 2010). De plus, entre écotypes, c'est à dire, entre des populations de l'espèce *A. thaliana* qui présentent des caractéristiques nouvelles adaptées à leur habitat, des différences au niveau de la présence et de l'expression de ces variants ont été observées (Abou-Ellail et al., 2011; Chandrasekhara et al., 2016).

E. Organisation des gènes d'ARNr au sein du génome

Les gènes d'ARNr 5S peuvent être organisés de manières différentes dans les génomes :

- dispersés dans le génome chez les champignons (Goyon, 1996; Selker et al., 1981)

- présents en répétition en tandem; soit intégrés dans les répétitions des gènes d'ARNr 45S (organisation de type-L (linked)) aussi bien en sens qu'en anti-sens (Sone et al., 1999; Vahidi et al., 1988); soit comme pour l'homme, la majorité des animaux et des plantes supérieures, en cluster(s) séparé(s) des unités 45S (organisation de type-S (separated)).

Les régions chromosomiques composées d'ADNr semblent présenter une certaine dynamique car leur position au sein du génome peut changer d'une espèce à une autre par translocation (Cazaux et al., 2011; Stults et al., 2009). Ainsi, chez les plantes des analyses phylogénétiques ont montré qu'il y a eu plusieurs événements de réorganisation des gènes d'ADNr 5S entre le type-L et le type-S (Wicke et al., 2011). D'un point de vue fonctionnel et adaptatif, les études réalisées chez les plantes ne semblent pas conclure en faveur d'un avantage sélectif à posséder une organisation de type L ou S (Garcia et al., 2007, 2009; Garcia and Kovařík, 2013; Wicke et al., 2011).

A. thaliana possède une organisation de type-S avec les gènes d'ARNr 45S organisés en répétitions directes (répétitions en tandem) dans les zones sub-télomériques des chromosomes 2 et 4 et les gènes d'ARNr 5S présents dans les régions péricentromériques de certains chromosomes. L'ensemble des répétions d'un *locus* sera ensuite défini comme un « ***cluster*** ».

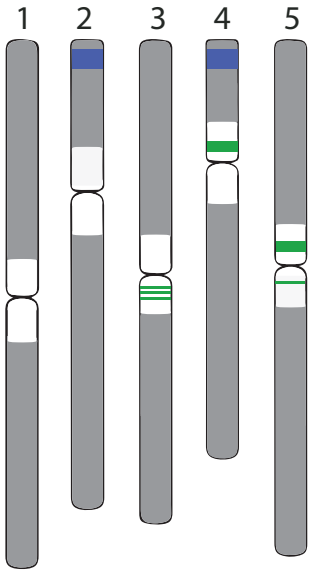


Figure 6 : Localisation des loci d'ADNr 5S et des loci d'ADNr 45S chez *Arabidopsis thaliana*, écotype Columbia (Col-0).
Les gènes d'ARNr 5S (vert) sont présents dans les zones péricentomériques (en blanc) des chromosomes 3, 4 et 5. Le chromosome 3 porte trois petits *loci* (Cloix et al., 2002). Le chromosome 5 porte deux *loci*: un sur le bras long et un petit sur le bras court. Les gènes d'ARNr 45S (en bleu) sont localisés en région sub-télomérique des chromosomes 2 et 4.

```
                    +121
                      |
Chr.3 locus 1         TTTTTTTCGGTTTTCTCTTTTTTTTTGTC
Chr.3 locus 2         TTTTTT-CGGTTTTCTCTTTTTTTTTGTT
Chr.3 locus 3         TTTTTTTCGGTTTTCTCTTTTTTTTTGTT
Chr.4                 TTTTTTTTTTTTTTTTTT------GGTT
Chr.5 locus majeur    TTTTATGTTTAACCTTTTTTTTT-GGTT
Chr.5 locus mineur    TTTTTTTTTAATTTGTTTTTTT---GGTT
```

Figure 7 : Alignement des "T-stretch" des gènes d'ARNr 5S chez *Arabidopsis thaliana*, écotype Columbia (Col-0).
Les séquences signatures "T-stretch" sont constituées d'une répétition de thymines dont la longueur et les motifs qui les interrompent sont spécifiques à chaque *locus*.

Pour l'écotype Columbia (Col-0) d'*A. thaliana* les principaux clusters d'ARNr 5S sont situés dans les régions péricentromériques des chromosomes 3, 4 et 5 (Figure 6) (Fransz et al., 1998, 2000; Murata et al., 1997). Plus précisément, on observe dans la zone péricentromérique du bras court du chromosome 3, 3 petits *loci* d'ADNr 5S (qui dans la suite de ce manuscrit seront considérés ensemble comme le *locus* du chromosome 3), un *locus* sur le bras court du chromosome 4 et un grand *locus* sur le bras court du chromosome 5 et un petit *locus* très muté sur le bras long du chromosome 5. Il existe également plusieurs petits clusters constitués de quelques unités de gènes d'ARNr 5S à différentes localisations chromosomiques (Murata et al., 1997).

La caractérisation plus détaillée de ces clusters par séquençage d'environ 300 unités d'ADNr 5S issues de YAC (yeast artificial chromosome) et de BAC (bacterial artificial chromosome) a permis de mettre en évidence une séquence riche en thymine en aval du gène (ensuite appelé » « ***T-stretch*** ») qui porte une signature nucléotidique spécifique pour chaque chromosome (Figure 7). En effet, le T-stretch du chromosome 3 comprend un motif CTC...CGG, celui du chromosome 4 est un polymère de T ininterrompu et celui du 5 contient le motif ATG...AACC. Cette signature a été mise à profit pour identifier l'origine chromosomique d'un gène d'ARNr 5S (Cloix et al., 2000, 2002). Cependant, cette analyse a été réalisée sur peu de séquences et sur des unités d'ADNr 5S qui ont été amplifiées par PCR ce qui a pu amener un biais dans l'analyse de ces séquences.

La présence des clusters d'ADNr 5S a été étudiée par cytologie en hybridation *in situ* chez différents écotypes d'*A. thaliana*. Cette analyse a montré que le nombre et la position des clusters d'ADNr 5S varient d'un écotype à l'autre chez *A. thaliana* (Fransz et al., 1998). Par exemple, les écotypes Col-0, Cape Verde Island (Cvi) et Kashmir1 (Kas) présentent des clusters communs d'ADN 5S alors que les écotypes Wassileskija (WS), C24 et Landsberg (La-0) ne possèdent pas de cluster sur le chromosome 3 (Fransz et al., 1998).

Une mutagénèse par irradiation a été réalisée sur des graines de l'écotype Landsberg (La) et un nouveau écotype (Landsberg *erecta* (L*er*)) a été sélectionné ensuite (Rédei, 1992) qui présente des caractéristiques phénotypiques distinctes de La-0. L*er* contient la mutation *erecta* qui provoque une altération du développement des organes de la plante (Rédei, 1992; Torii et al., 1996). Des analyses d'hybridation *in situ* (Fluorescence *in situ* hybridization) (FISH) ont également montré que L*er* possède un *locus* d'ADNr 5S sur le bras long du

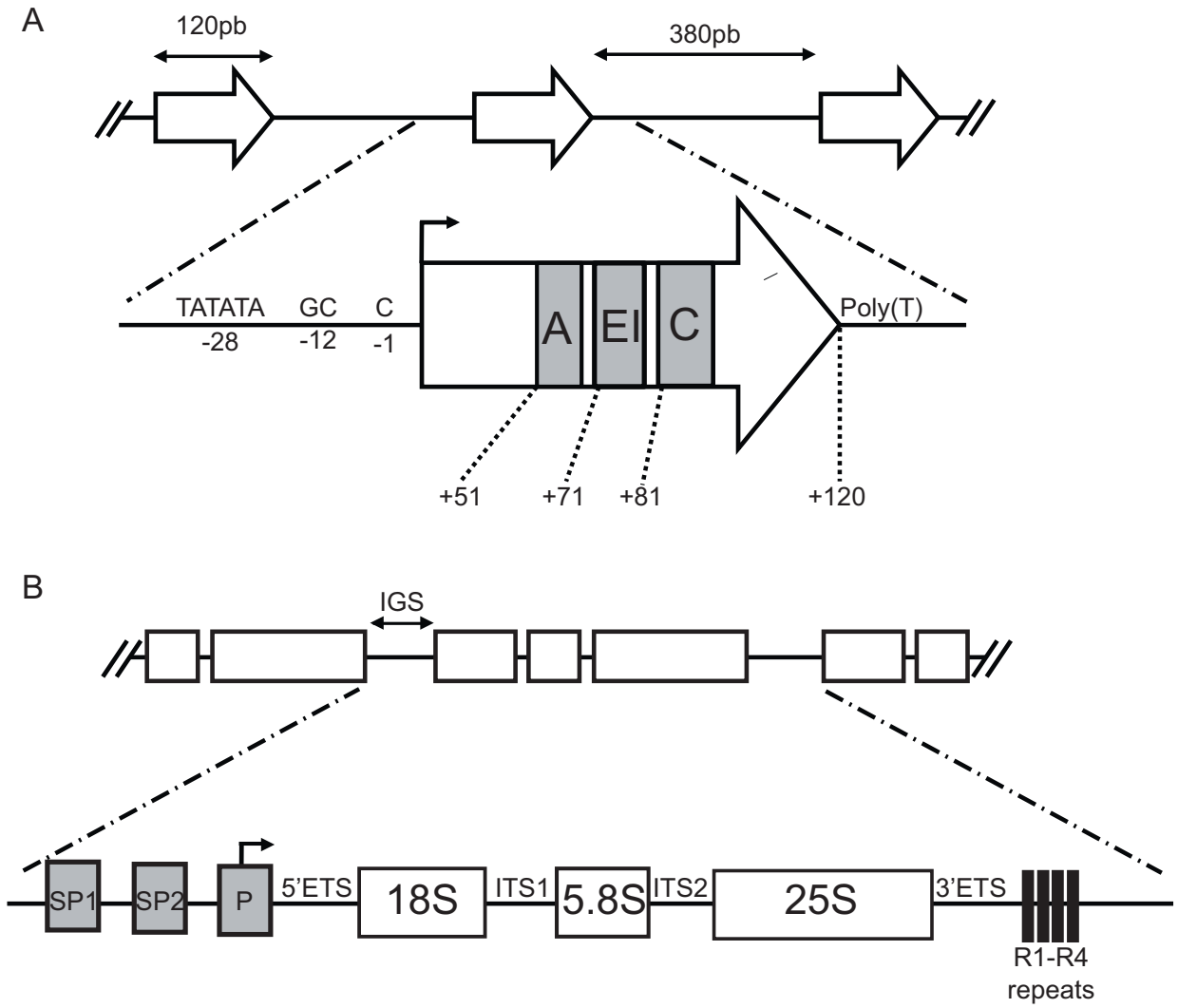


Figure 8 : Structure des gènes d'ARNr 5S et d'ARNr 45S chez *Arabidopsis thaliana*.
(A) Les gènes d'ARNr 5S sont organisés en répétitions en tandem de séquences transcrites de 120 pb séparées par des séquences intergéniques de 380 pb. Le zoom présente les régions régulatrices identifiées dans un gène d'ARNr 5S. La séquence promotrice interne est composée d'une boîte A, d'un élément intermédiaire (EI) et d'une boîte C. On trouve une boîte TATA-like, un di-nucléotide GC et une cytosine en amont et une séquence riche en thymine est présente en aval de la séquence transcrite.

(B) Les gènes d'ARNr 45S sont composés des ADNr 18S, 5.8S et 25S. Ils sont séparés par des espaceurs internes (ITS1 et ITS2). On trouve aux extrémités de ces séquences deux espaceurs externes (3'ETS et 5'ETS). En amont du gène, on trouve un promoteur (P) et deux espaceurs promoteurs (SP). En aval des gènes se trouvent 4 répétitions qui sont responsables de la variabilité des transcrits 45S.

chromosome 3 et non pas dans les zones péricentromériques comme c'est le cas pour le chromosome 3 de Col-0 (Fransz et al., 1998).

Les gènes d'ARNr étant des séquences dynamiques dans le génome des plantes (Wicke et al., 2011) aussi bien dans l'organisation en type L ou en type S qu'en nombre de clusters existant dans un génome, ils sont un excellent modèle d'étude des processus de dynamique du génome.

F. Structure des gènes d'ARNr chez *Arabidopsis thaliana*

Les gènes d'ARN 5S d'une longueur de 500bp chez *A. thaliana* sont composés d'une séquence transcrite de 120bp et d'une séquence intergénique de 380pb (Figure 8A). Il existe en plus sur le petit *locus* du chromosome 5 des unités plus petites de 251pb provenant d'une délétion dans la séquence transcrite et intergénique (Cloix et al., 2000).

Les gènes d'ARNr 5S possèdent un promoteur de type 1 qui est un promoteur interne tripartite ou ICR (internal control region) composé d'une boîte A, d'un élément interne et d'une boîte C. Ces trois éléments ont été identifiés par homologie de séquences et des analyses de transcription *in vitro* ont révélé que des mutations ponctuelles dans les trois composants du promoteur interne entraînent une baisse de la transcription des gènes d'ARNr 5S (Cloix et al., 2000, 2003). Le promoteur interne permet la fixation du facteur TFIIIA et la transcription des gènes d'ARNr 5S.

L'ARN Pol III reconnaît chez *A. thaliana* quatre thymines comme signal d'arrêt de la transcription. Ces quatre thymines font partie de la séquence 'T-stretch' d'environ 20 à 30 bp située en aval de la séquence transcrite.

En amont du gène, dans la zone intergénique, plusieurs séquences régulatrices ont été identifiées par homologie de séquences (Venkateswarlu et al., 1991). Il s'agit d'un motif TATA (-28 à -23 en amont de la séquence transcrite), du dinucléotide GC (-11 à -12) et de la cytosine (-1) qui sont indispensables pour un niveau correct de transcription des gènes d'ARNr 5S *in vitro*. En effet, la présence de mutations dans ces séquences réduit l'efficacité de la transcription des gènes d'ARNr 5S et le motif TATA est essentiel à la transcription et la ré-initiation de la transcription des gènes d'ARNr 5S (Cloix et al., 2003).

Les gènes d'ARNr 45S (~10Kb) possèdent une structure plus complexe (Figure 8B) où les 3 séquences codant les ARNr sont séparées par deux espaceurs internes (ITS1 et

ITS2). Deux espaceurs externes (3'ETS et 5'ETS) sont localisés aux extrémités de ces séquences (Grummt and Pikaard, 2003; Pikaard, 2002). Les ETS et ITS font partie du pré-ARNr 45S et sont excisés lors de la maturation. Les gènes d'ARNr 45S sont séparés par des séquences intergéniques (IGS) composées d'un promoteur, de séquences répétées et de promoteurs espaceurs (McStay and Grummt, 2008). Le promoteur proximal a été identifié entre -55 et + 6 et les séquences localisées entre -33 et +6 sont suffisantes pour la transcription.

G. Dynamique du nombre de copies des gènes d'ARNr

Comme les ARNr sont indispensables à la vie de la cellule et qu'ils participent à la synthèse protéique, les ARNr doivent être transcrits de façon abondante afin de répondre aux besoins de la cellule surtout lorsque l'organisme est en croissance et donc quand le besoin en protéine est important. Afin de répondre à la demande en ARNr 5S, les gènes d'ARN 5S sont présents en copies multiples.

Le nombre de ces copies de gènes d'ARNr est variable entre les espèces, par exemple seules 8 copies sont présentes chez *Escherichia coli* (type L) (Condon et al., 1995) mais plus de 60 000 copies sont présentes chez *Linum usitatissimum* (Schneeberger et al., 1989). Il a été estimé que l'écotype Col-0 d'*A. thaliana* contient par génome haploïde ~1000 copies d'ADNr 5S (Campell et al., 1992) et ~700 copies d'ADNr 45S (Copenhaver and Pikaard, 1996). Les ADNr représentent (avec les répétitions 180) la fraction répétée la plus importante du génome d'*A. thaliana* (Davison et al., 2007).

Concernant les clusters d'ADNr 5S, des analyses de FISH sur l'écotype Col-0, indiquent que le nombre de gènes du cluster du chromosome 3 est plus petit que celui du chromosome 4 qui serait de taille intermédiaire et que celui du bras supérieur du chromosome 5 serait le plus grand (Murata et al., 1997).

Pour *A. thaliana*, une pseudomolécule du génome assemblée est disponible (https://www.arabidopsis.org/). Seules 290 copies d'ARN 5S sur le chromosome 3, 14 sur le chromosome 4 et 59 sur le chromosome 5 ont été recensées. Ceci n'est pas cohérent avec les estimations du nombre de copies (Campell et al., 1992) ou les analyses de FISH (Fransz et al., 1998; Murata et al., 1997). La pseudomolécule est donc incomplète et/ou mal assemblée au niveau des séquences hautement répétées.

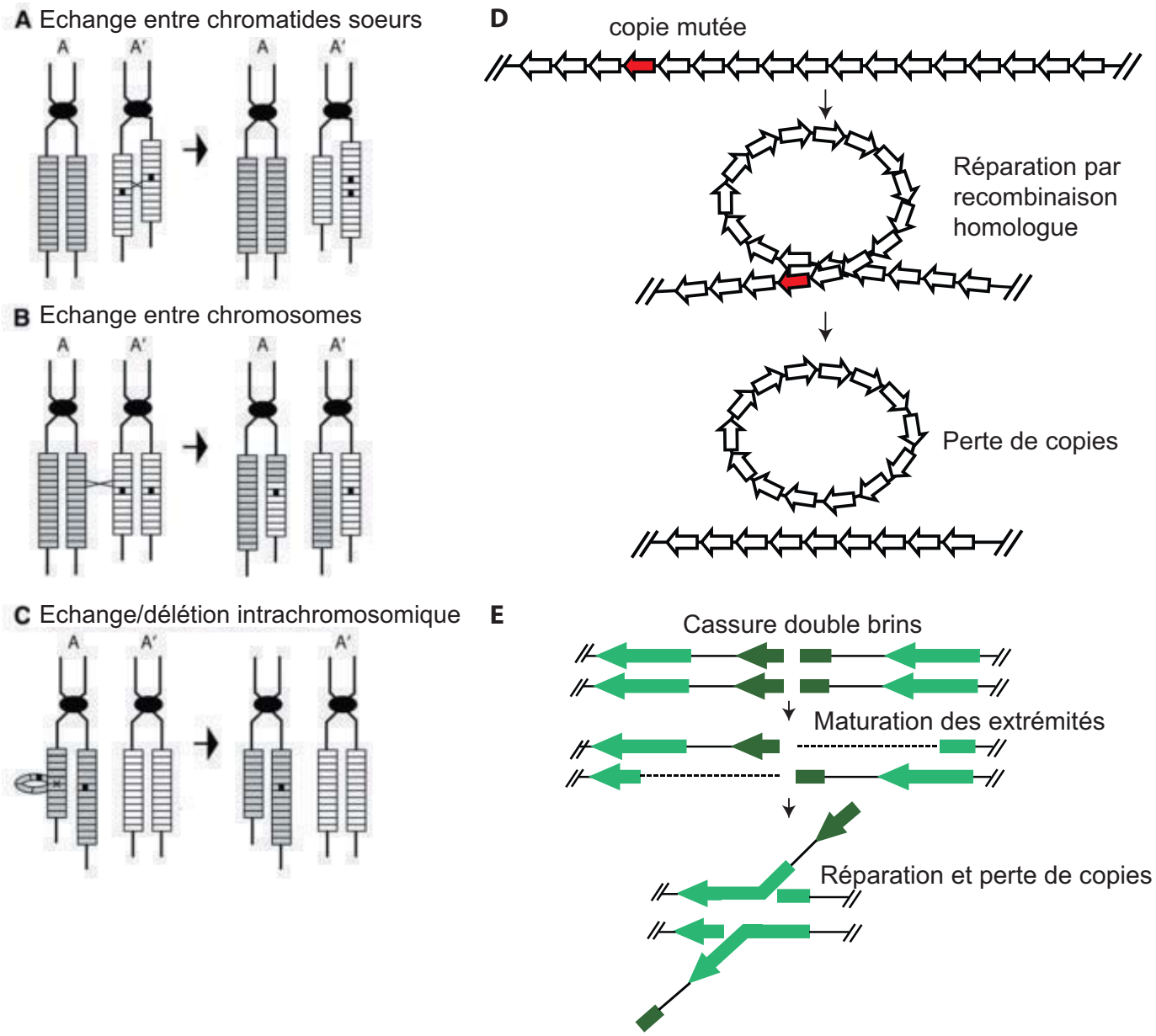


Figure 9 : Processus impliqués dans l'augmentation ou la réduction du nombre de copies des gènes d'ARNr.

(A/B/C) A et A' représentent des chromosomes homologues, et les rectangles des unités d'ADNr. Les points noirs représentent des mutations. Les 3 mécanismes de recombinaisons peuvent amener à une duplication ou à une perte d'unité d'ADNr sur le chromosome par crossing-over (Eickbush and Eickbush, 2007).

(D) Des dommages ADN dans un gène d'ARNr peuvent entrainer sa réparation par recombinaison avec une autre copie et ainsi causer la perte des gènes d'ARNr présents entre ces 2 gènes. D'après Kobayashi 2011.

(E) Quand une cassure double brin est créée au sein d'un gène d'ARNr, la voie de réparation SSA (Single-strand annealing) va pouvoir réparer l'ADN ce qui va entraîner la perte de l'unité mutée. D'après Kobayashi 2011.

La dynamique dans le nombre de copies d'ADNr observées entre clusters et à l'intérieur d'un cluster pourrait s'expliquer par le fait que les gènes d'ARNr sont des zones préférentielles de cassure ADN et subissent de par leur nature répétée des variations du nombre de copies par des mécanismes de recombinaison homologue (Figure 9) (Eickbush and Eickbush, 2007; Guetg et al., 2010; Kobayashi, 2011). Il existe donc une certaine variabilité du nombre de copies de gènes d'ARNr 5S à l'intérieur d'une même espèce (Cazaux et al., 2011; Gibbons et al., 2015). Toutefois, il semble exister des mécanismes de régulation permettant de maintenir un nombre minimal de copies d'une part pour maintenir une synthèse protéique suffisante et d'autre part pour maintenir la stabilité du génome (voir paragraphe H). Ainsi, chez *S. cerevisiae*, un système de maintien du nombre de copies d'ADNr permet l'amplification de gènes afin de palier à la perte de copies par les systèmes de réparation de l'ADN (Figure 10). Pendant la phase S, la réplication commence à l'origine de réplication et une des fourches est bloquée par la protéine Fob1 (fork blocking protein 1) qui vient se placer sur une séquence spécifique au niveau des gènes d'ARNr. Ce blocage fonctionne comme un signal pour la recombinaison et génère une cassure double brin par une nucléase. Il existe au niveau de l'arrêt de la réplication un promoteur (E-pro) produisant des ARNs ayant la propriété de déstabiliser la cohésion entre chromatides sœurs en décrochant les molécules de cohésines. Lorsque le nombre de gènes d'ARNr 5S est optimal, E-pro est inhibé par une histone déacétylase, Sir2, ce qui favorise la réparation par recombinaison homologue entre les chromatides sœurs. Lorsque le nombre de copies d'ADNr est réduit, Sir2 ne se fixe pas sur E-pro qui est transcrit ce qui entraîne le décrochage des cohésines et favorise la réparation par recombinaison inégale entre chromatides sœurs. Ce mécanisme conduit à une augmentation du nombre de copies de gènes d'ARNr 5S (revue : Kobayashi, 2011).

Malgré le fait que les gènes d'ARNr 5S et 45S soient positionnés sur des chromosomes différents (organisation de type S) et ne montrent pas de similarité de séquences, il existe chez l'homme et la souris une corrélation positive entre le nombre de copies d'ADNr 45S et 5S (Gibbons et al., 2015). Un tel mécanisme pourrait s'apparenter à des observations faites lors de duplication du génome. En effet, lorsque des gènes impliqués dans la synthèse de protéines composant un complexe sont dupliqués, ils mettent plus de temps que les autres à revenir à un état monogénique. L'une des hypothèses serait qu'un déséquilibre entre le nombre de gènes des protéines composant un complexe soit délétère pour la cellule (Birchler and Veitia, 2012).

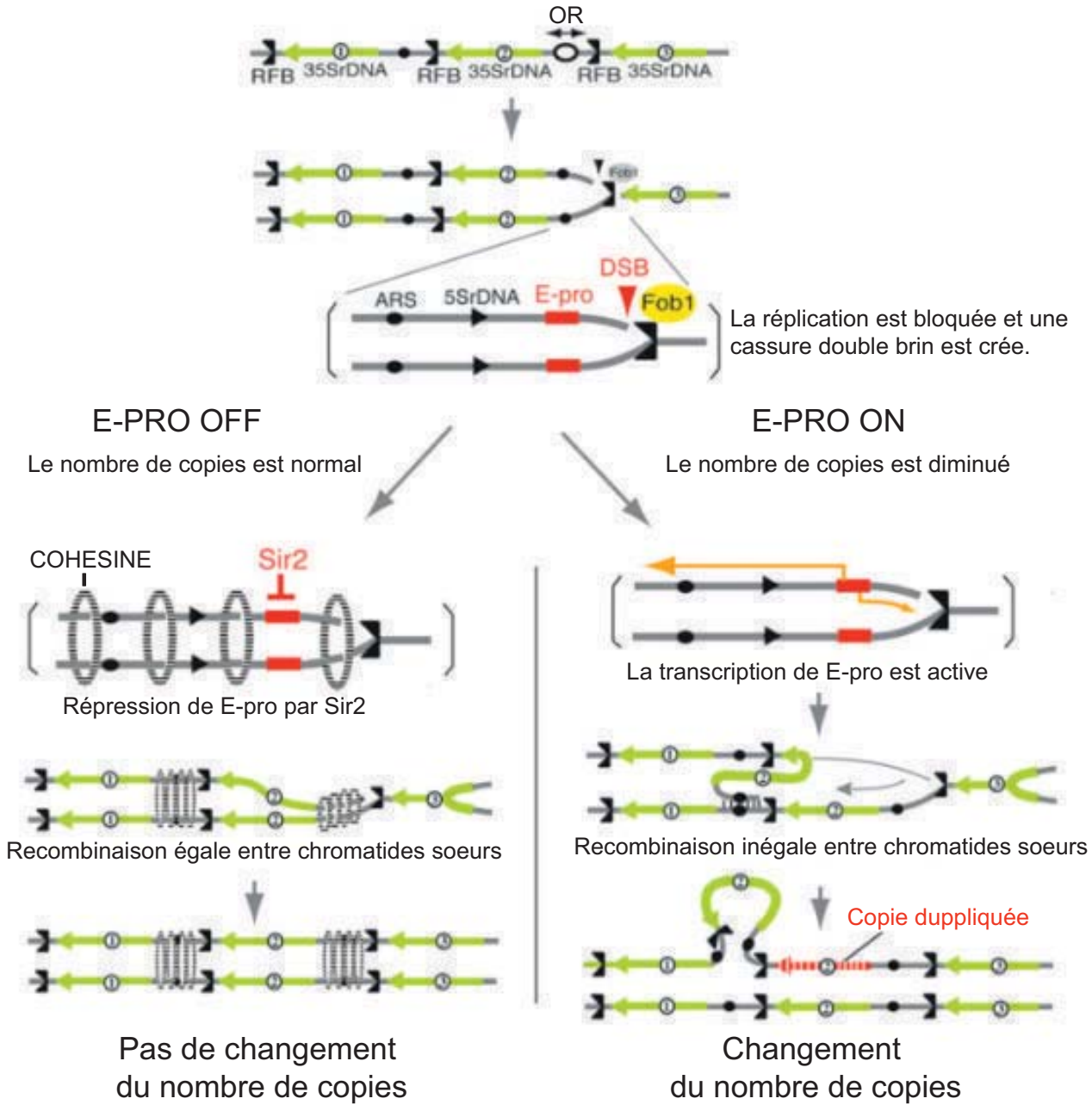


Figure 10 : Mécanisme de régulation du nombre de copies d'ADNr chez la levure.
La réplication commence à l'origine de réplication (OR) et la protéine Fob1 (Fork Blocking Protein 1) vient se placer sur une séquence spécifique au niveau des gènes d'ARNr, ce qui provoque un arrêt de la fourche de réplication et induit une cassure double brin. Il existe au niveau de l'arrêt de la réplication, un promoteur (E-pro) dont les transcrits détachent la cohésine des chromatides sœurs. Quand le nombre de gènes d'ARNr 5S est adéquat, E-pro est inhibé par la protéine Sir2 ce qui favorise la réparation par recombinaison égale entre chromatides sœurs car la cohésine est toujours présente. Si le nombre de copies d'ADNr est diminué, une transcription à partir d'E-pro décroche la cohésine ce qui favorise la réparation par recombinaison inégale des chromatides sœurs ce qui amène à une augmentation du nombre de copies. D'après Kobayashi, 2011).

Il existe chez *A. thaliana* des mutants pour lesquels le nombre de copies d'ADNr est augmenté ou diminué par rapport à une plante sauvage. C'est le cas des mutants pour les gènes *ATXR5/ATXR6* (Alpha Thalassemia-mental Retardation X-linked) qui sont des histones méthyltransférases (Jacob and Feng, 2009), responsables de la mise en place d'une mono-méthylation sur l'histone H3 (H3K27me1) et les gènes du complexe CAF-1 (Chromatin assembly factor 1), une chaperonne d'histones H3-H4 qui permet l'assemblage des nucléosomes lors de la réplication (Ramirez-Parra and Gutierrez, 2007a). Le double mutant *atxr5 atxr6* induit une augmentation du nombre d'ADNr 45S (Pontvianne et al., 2012) alors que les mutants *fasciata* (*fas*) 1 et 2*,* deux sous unités du complexe CAF-1, induisent une réduction du nombre de copies d'ADNr 45S (Pontvianne et al., 2013). L'effet des mutants *fas1* et *fas2* est spécifique des gènes d'ADNr 45S et cumulatif d'une génération à une autre avec une perte estimée entre 10 et 20% par génération (Mozgová et al., 2010). Cette perte de copies d'ADNr 45S touche ne réduit pas le niveau global de transcription du fait de l'induction de l'expression du variant 1 normalement silencieux dans une plante adulte (Pontvianne et al., 2010, 2013). La réintroduction de FAS1 dans ces plantes permet une réversion partielle de la perte des gènes d'ARNr 45S ce qui suggère que les plantes possèdent également un mécanisme de régulation du nombre de copies d'ADNr qui reste à découvrir.

Il semble exister une corrélation positive entre la taille du génome et le nombre de copies d'ADNr comme le suggère l'analyse de 162 espèces de plantes et d'animaux (Prokopowich et al., 2003). C'est aussi le cas pour certains écotypes d' *A. thaliana* notamment pour une sous population présente au nord de la Suède (Long et al., 2013). Ces données suggèrent soit que l'augmentation de la taille du génome conduit à une augmentation du besoin en traduction et donc nécessite un nombre de copies plus élevé d'ADNr soit que le nombre de copies d'ADNr serait un élément important pour la stabilité des génomes (voir paragraphe H).

Enfin, il convient de noter que seule une partie des gènes d'ARNr est transcrite (Cloix et al., 2002; Dammann et al., 1993; French et al., 2003; Grummt, 2010). Le maintien dans le génome des copies inactives suggère également que les gènes d'ARNr auraient d'autres rôles que la synthèse de l'ARNr par exemple en participant au maintien de la stabilité du génome.

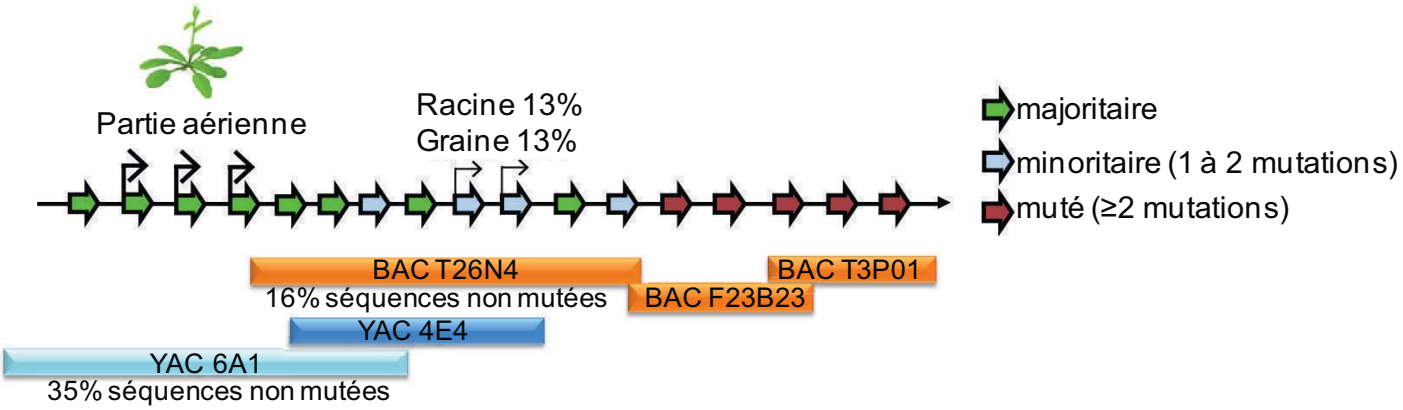


Figure 11 : Localisation des YACs et des BACs couvrant les clusters d'ADNr 5S du chromosome 5 et caractérisation des gènes d'ARNr 5S.
Les chromosomes artificiels de levure (YAC) et les chromosomes artificiels de bactéries (BAC) ont été cartographiés sur le cluster d'ADNr 5S du bras court du chromosome 5. Le YAC 6A1 correspond à la fraction la plus euchromatique du cluster d'ADNr 5S et le YAC 4E4 à la partie la plus proche du centromère. Le nombre de polymorphismes présents au sein des gènes d'ARNr 5S augmente lorsque l'on se dirige vers la partie la plus centromerique du cluster. Il existe 2 types de gènes d'ARNr 5S : (i) les gènes majoritaires dont la séquence transcrite est identique à la séquence consensus et qui sont majoritairement exprimés dans la partie aérienne de la plante ; (ii) les gènes minoritaires qui portent un ou plusieurs polymorphismes. Seul les séquences présentant un ou deux polymorphismes sont exprimés dans certains tissus ou à certains stades de développement.

H. ADNr et stabilité du génome

Les arguments en faveur d'un rôle des gènes d'ARNr dans le maintien de la stabilité du génome sont sujets à débat. En effet, d'une part, les gènes d'ARNr de par leur nature hautement répétée font partie des régions les plus instables du génome (Cazaux et al., 2011; Kobayashi, 2008; Stults et al., 2009). Chez l'homme, les gènes d'ARNr 45S constituent des points de fragilité les plus importants du génome par l'apparition de cassures au niveau des IGS (Tchurikov et al., 2015). Les translocations impliquant les ADNr 45S constituent d'ailleurs l'une des altérations du chromosome les plus fréquentes retrouvées dans les tumeurs chez l'homme (Stults et al., 2009). Cependant, au sein des ADNr, les cellules humaines seraient plus sensibles aux dommages ADN touchant les unités 45S qu'aux dommages touchant les unités 5S (Warmerdam et al., 2016).

D'autre part, des études effectuées sur des lignées de levure présentant un nombre variable de copies de gènes d'ARNr ont montré que les gènes d'ARNr et notamment ceux qui sont non transcrits sont importants pour la stabilité du génome. En effet dans les souches présentant seulement 20 copies d'ADNr 45S, toutes les copies sont transcrites, mais ces cellules montrent une sensibilité augmentée aux dommages UV et aux agents mutagènes (Ide et al., 2010). Dans ces souches avec peu de copies de gènes d'ARNr, la condensine ne peut s'associer aux ADNr transcrits ce qui provoque une séparation prématurée des chromatides sœurs et empêche une réparation correcte des ADNr mais aussi des autres séquences du génome lors des dommages ADN (revue : Kobayashi, 2011).

I. Hétérogénéité des séquences d'ADNr et des transcrits ARNr

Chez *A. thaliana*, la cartographie et le séquençage Sanger de YAC et de BAC couvrant les clusters d'ADNr 5S a permis de redéfinir des séquences consensus pour chaque *locus* et d'observer une variabilité au niveau du taux de mutations/polymorphismes (Cloix et al., 2000; Creusot et al., 1995; Tutois et al., 1999). En effet, les gènes d'ARN 5S des clusters du chromosome 4 et 5 montrent peu de polymorphismes alors que celui du chromosome 3 et le petit *locus* du chromosome 5 sont très polymorphes. Ces analyses montrent également que la séquence intergénique est plus polymorphe que la séquence transcrite. Enfin, les séquences identiques à la séquence consensus sont enrichies dans le YAC le plus euchromatique alors que les polymorphismes augmentent pour les séquences plus proches du centromère (Figure 11). Il existe donc un gradient de polymorphismes le long du cluster du chromosome 5 (AGI, 2000; Cloix et al., 2002).

Des analyses *in vitro* ont permis de montrer que les séquences présentes sur le *locus* du chromosome 3 et sur le petit *locus* du chromosome 5 ne sont pas transcrites notamment de par la présence de mutations dans le promoteur interne ou dans les séquences régulatrices sur ces deux *loci* (Cloix et al., 2002, 2003). Ce résultat est concordant avec le fait que 85% des gènes d'ARNr 5S des YAC du chromosome 3 portent une mutation (TATATA en TAGATA) du motif TATA-like indispensable à la transcription de ces gènes (Cloix et al., 2002).

Les gènes d'ARNr 5S portant plus de deux polymorphismes dans la séquence transcrite ne sont jamais exprimés. Dans la partie aérienne de la plante, seuls les gènes présentant la séquence consensus de l'ARNr 5S sont transcrits (Mathieu et al., 2003a). On appelle ces gènes les **gènes majoritaires** (par analogie avec les ARNr qu'ils produisent). Les **gènes minoritaires** possèdent un à deux polymorphismes (minoritaires peu polymorphes) ou au moins 3 mutations (minoritaires hautement polymorphes) dans la partie transcrite de l'ADNr 5S.

Les gènes minoritaires portant entre 1 et 2 polymorphismes ne sont pas exprimés dans la partie aérienne de la plante mais sont exprimés dans certains tissus comme la racine, ou certains stades du développement comme la graine et la racine. De plus, ils sont exprimés dans certains mutants qui affectent l'organisation chromatinienne et la mise en place de marques épigénétiques (Mathieu et al., 2003a; Vaillant et al., 2007). Les ARNr 5S minoritaires seraient exprimés dans les tissus et pendant les stades du développement caractérisés par un fort besoin en traduction ou à un moment où certaines marques épigénétiques établissant le statut transcriptionnel de la chromatine ne sont pas encore totalement mises en place (Douet et al., 2008; Layat et al., 2012).

Au final, seuls les *loci* du chromosome 4 et le grand *locus* du chromosome 5 semblent transcrits et au sein de ces deux *loci* seuls les ARNr 5S majoritaires sont transcrits de façon constitutive.

Concernant les gènes d'ARNr 45S, certains polymorphismes identifiés sont localisés dans la séquence transcrite excisée lors de la maturation et ne seront donc pas retenus dans les ARNr matures. Il existe de plus, quatre à cinq variants d'ADNr 45S qui sont caractérisés par des régions 3' ETS distinctes: le variant 1 (VAR1) qui comporte une insertion de 80 pb (48% des séquences d'ADNr 45S), le variant 3 (VAR3) qui comporte une délétion de 90 pb (22% des séquences), les variants 2 (VAR2) (30% des séquences) et 4 (VAR4) (très peu de

séquences) qui possèdent une petite insertion et/ou délétion supplémentaire (Abou-Ellail et al., 2011; Pontvianne et al., 2010). Il existe aussi des sous-variants de VAR2 et VAR3 qui se distinguent par une délétion de 3 bases (respectivement CAC et CGC) (Pontvianne et al., 2010). Plus récemment un 5ème variant a été décrit, VAR5, qui est similaire à VAR3 mais présente une délétion de 100 pb dans le 3' ETS (Havlová et al., 2016). VAR5 pourrait être considéré comme un sous-variant de VAR3. La présence de VAR1-5 varie d'un écotype à un autre, certains ne possèdent par exemple que deux variants (*Landsberg erecta* et *Wassilewskija*) ou un seul (*C24* et *Shahdara*) (Abou-Ellail et al., 2011; Chandrasekhara et al., 2016).

Bien que VAR1 soit le plus représenté au niveau ADN dans l'écotype Col-0, il n'est pas exprimé dans une plante sauvage excepté dans la graine, et dans certains mutants qui montrent une altération de l'organisation du nucléole, comme la nucléoline (Pontvianne et al., 2010). VAR4 est quant à lui très peu présent dans le génome, et un tel variant a été identifié dans la région péricentromérique du chromosome 3 (Chr3-VAR4), c'est à dire au milieu d'une région hétérochromatique où se trouvent également des gènes d'ARNr 5S non transcrits (Abou-Ellail et al., 2011). Le gène Chr3-VAR4 est transcriptionnellement actif malgré le contexte répété et transcriptionnellement inactif de l'environnement chromatinien local. La « protection » du gène Chr3-VAR4 serait établie par la nucléoline, un composant majeur du nucléole. En effet lorsque la nucléoline est inactivée par une mutation, les gènes d'ARNr 45S du type VAR4 sont réprimés. Il est aussi possible que les gènes d'ARNr 45S du type VAR4 ne donnent pas d'ARNr mature mais que l'ARN soit pris en charge par la voie RdDM (RNA-directed DNA methylation) et participe à la mise sous silence de certains gènes d'ARNr 45S (Abou-Ellail et al., 2011).

Les gènes d'ARNr 5S sont des séquences répétées localisées dans des régions dites hétérochromatiques, péricentromériques. Ces régions sont réputées pour former un environnement chromatinien répressif pour la transcription. Malgré cela, une partie des gènes d'ARNr 5S sont hautement exprimés. Il doit donc exister des mécanismes de régulation de la transcription des gènes d'ARNr ainsi que le choix des gènes qui sont mis sous silence. Il existe plusieurs niveaux de régulation : transcriptionnel avec l'intervention de facteurs de transcription comme TFIIIA, qui est finement régulé (voir chapitre II B) mais aussi au niveau local de la chromatine voire à un niveau d'ordre supérieur d'organisation des séquences d'ADNr dans le noyau. Ce sera l'objet du chapitre suivant.

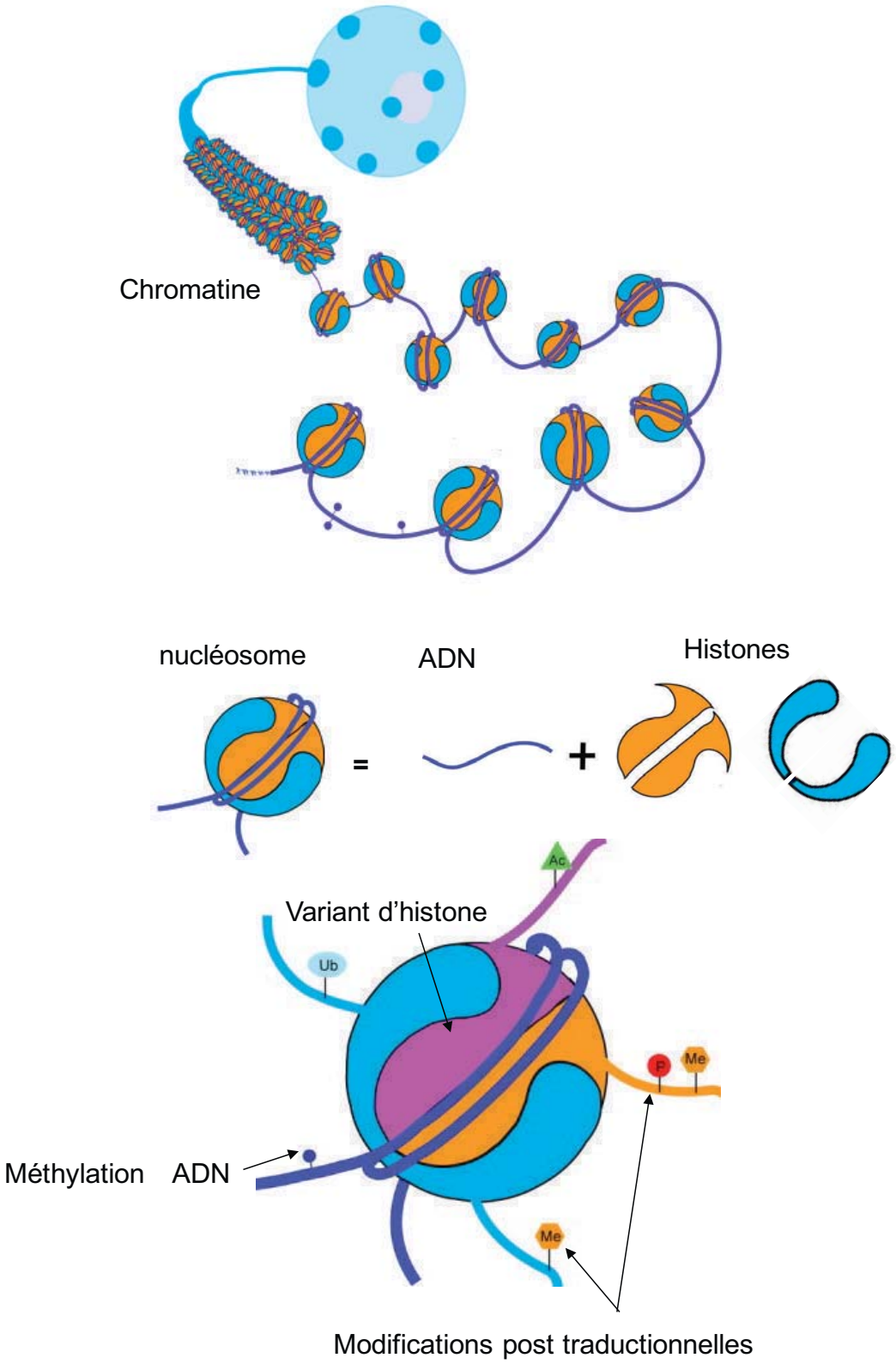


Figure 12 : L'organisation de la chromatine
Dans un noyau eucaryote, l'ADN est associé à un tétramère protéique d'histones H3-H4 et de deux dimères d'histones H2A-H2B formant ainsi le 'nucleosome core particle'. Les marques épigénétiques se situent à différents niveaux : sur l'ADN (sous forme de méthylation surtout sur les cytosines), sur les histones (sous forme de modifications post-traductionnelles comme la méthylation, l'acétylation ou la phosphorylation) et au niveau de la composition du nucléosome par la présence de différents variants d'histones (adapté de Probst et al., 2009).

III. L'épigénétique

Les cellules d'un organisme eucaryote contiennent toutes le même patrimoine génétique pourtant elles varient d'un point de vue transcriptionnel lors du développement, dans différents tissus ou encore lors de stress environnementaux entraînant des modifications possibles du phénotype. Cette modification du profil d'expression des gènes d'une cellule sans modification de la séquence nucléotidique s'appelle l'épigénétique.

La chromatine, l'organisation de l'ADN dans un complexe nucléoprotéique, a émergé ces dernières années comme un des vecteurs majeurs de la régulation épigénétique. La chromatine permet des changements potentiellement héritables d'expression des gènes sans d'altération de la séquence ADN.

A. La chromatine

Dans les cellules eucaryotes, l'ADN n'est pas nu mais il est emballé et compacté dans une structure nucléo-protéique: la chromatine (Figure 12). Le nucléosome est l'unité de base de la chromatine. Il est composé d'un octamère de protéines histones autours duquel vient s'entourer 146 pb d'ADN plus un ADN linker de 20 pb à 90 pb d'ADN qui peut être lié par l'histone 'linker' H1. L'octamère d'histones est composé de quatre histones différents : H3, H4, H2A et H2B. La composition riche en acides aminés basiques lysine et arginine confère aux histones une charge positive qui permet leur liaison à l'ADN globalement chargé négativement grâce à 14 liaisons non covalentes. Les histones possèdent une partie globulaire interne composée de trois hélices alpha et de boucles et de parties plus externes N-ter et C-ter qui sont exposées à l'environnement cellulaire. Les histones vont pouvoir être modifiées par l'ajout de groupement chimique qu'on appelle des modifications post-traductionnelles des histones ou des marques d'histones.

L'assemblage du nucléosome commence par l'hétérodimérisation des histones H3 et H4 suivie par une dimérisation de deux dimères H3-H4. Le tétramère ainsi formé va se positionner sur l'ADN. Deux dimères d'histones H2A-H2B viennent par la suite compléter le nucléosome par des interactions entre les histones H2B et H4 (Loyola and Almouzni, 2004). Les histones H3-H4 ont un taux de renouvellement inferieur aux histones H2A-H2B et sont donc plus stables sur l'ADN (Kimura and Cook, 2001). Il a été montré que la densité des nucléosomes présents sur une séquence ADN allait former une structure plus ou moins

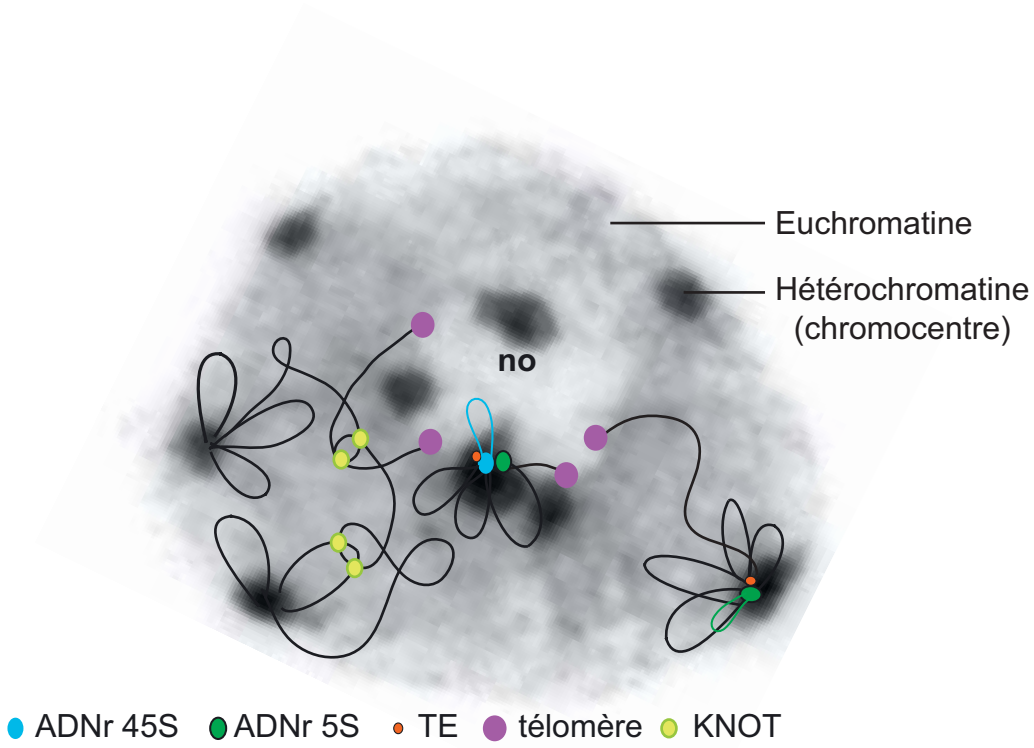


Figure 13 : Représentation schématique de l'organisation de l'ADN dans un noyau de feuille d'*Arabidopsis thaliana*.
Deux formes distinctes de chromatine ont été mises en évidence (Heitz, 1928): (i) l'euchromatine faiblement colorée au DAPI, riche en gène et plutôt permissive pour la transcription et (ii) l'hétérochromatine fortement colorée au DAPI qui correspond aux séquences répétées et aux éléments transposables. Les clusters d'hétérochromatine condensée dans les noyaux d'*Arabidopsis* sont appelés chromocentres. Les chromocentres contiennent les régions répétées du génome ainsi que les éléments transposables (TE) et ancrent des boucles d'euchromatine alors que les télomères se regroupent près du nucléole (no), qui est très peu marqué au DAPI. Il existe aussi des Interactive heterochromatic island (IHI)/KNOT engaged element (KEE) qui forment des interactions intras et inter-chromosomiques. Le chromocentre du chromosome 4 est associé au nucléole et comprend l'ADNr 45S, l'ADNr 5S ainsi que des éléments transposables.

compacte et donc plus ou moins permissive aux machineries transcriptionnelles, réplicatives et réparatrices (Probst et al., 2009; Zhang et al., 2015).

B. Les différents types de chromatine

La coloration au 4′,6′-diamidino-2-phenylindole (DAPI) de noyaux en interphase (Figure 13) permet d'observer des zones de forte intensité de fluorescence qui correspondent à l'hétérochromatine et des zones de faible fluorescence qui correspondent à l'euchromatine.

La forme la plus dense : l'hétérochromatine; est une structure compacte, riche en séquences répétées. Cet état de la chromatine présente un espacement régulier des nucléosomes (Chodavarapu et al., 2010) et est résistant à la DNAse I (Shu et al., 2012). La forte densité nucléosomale de l'hétérochromatine limite l'accès de la machinerie transcriptionnelle et favorise le maintien sous silence des éléments transposables. L'hétérochromatine présente une réplication tardive comparée à celle de l'euchromatine (Dileep et al., 2015).

La forme relâchée de la chromatine : l'euchromatine, présente une faible densité nucléosomale en comparaison avec l'hétérochromatine, est enrichie en gènes et est plutôt favorable à la transcription (Benoit et al., 2013).

L'hétérochromatine et l'euchromatine ne sont pas localisées de manière aléatoire dans le génome d' *A. thaliana*. L'euchromatine est enrichie au niveau des bras chromosomiques alors que l'hétérochromatine se trouve dans trois régions distinctes. La zone d'hétérochromatine centromérique est une région riche en séquences « satellite » telle que la séquence de 178pb appelée AtCon, pAL1 ou 180pb (Kumekawa et al., 2000, 2001; Nagaki et al., 2003) arrangée en répétitions en tandem de ~0,4 à 3 Mb . Les répétitions de 180pb sont interrompues par la séquence 106B qui dérive du rétrotransposon Athila2 (Thompson et al., 1996). La zone péricentromérique contient des éléments transposables, la séquence Transcriptional silent information (TSI) et les gènes d'ARNr 5S (Bauwens et al., 1991; Cloix et al., 2000; Fransz et al., 1998; Steimer et al., 2000). Enfin, la zone télomérique et sub-télomérique inclut les NORs (nucleolus organizer regions) qui contiennent les gènes d'ARNr 45S (Fransz et al., 2000). La structure et la fonction des zones centromériques et péricentromériques chez *A. thaliana* ont fait l'objet d'une revue : (Simon et al., 2015) (cf **Annexe I**).

L'hétérochromatine est également répartie de façon non aléatoire d'un point de vue cytologique. Chez *A. thaliana*, les régions centro- et péricentromériques sont condensées dans un noyau en interphase en 6 à 10 structures appelées les chromocentres localisés en périphérie du noyau (Figure 13) (Fransz et al., 2002a). Les analyses de FISH ont révélé que des boucles d'euchromatine d'une taille d'environ 0.2 à 2 Mb se forment à partir des chromocentres ce qui donne une organisation des chromosomes en "rosette" et définit des territoires chromosomiques (Figure 13) (Fransz et al., 2002a; Fransz and de Jong, 2011; Pecinka et al., 2004; Rodriguez-Granados et al., 2016). Il est possible que cette organisation en rosette favorise la répression transcriptionnelle des séquences localisées au cœur du chromocentre alors qu'elle favoriserait l'expression des séquences ADN dans les boucles.

Ces résultats ont été pour la plupart confirmés par des données récentes utilisant la technologie du Chromosome Conformation Capture (3C) étendue à l'échelle du génome (4C et Hi-C) qui permet d'identifier les interactions entre domaines de chromatine au sein du noyau (Grob et al., 2013, 2014). Ces analyses ont révélé que les séquences hétérochromatiques interagissent préférentiellement les unes avec les autres que ce soit au niveau des zones centro- et péricentromérique ou au niveau des régions hétérochromatiques intercalées dans l'euchromatine appelées des KNOT (Grob et al., 2013, 2014). Elles servent donc à structurer, organiser le génome d' *A. thaliana*. Une des différences observée par rapport au modèle de Fransz est que la partie proximale du chromosome s'organise avec les centromères mais que la partie distale du chromosome s'organise avec les télomères proches du nucléole.

Chaque type de chromatine, euchromatine et hétérochromatine, est caractérisée par une combinaison de modifications épigénétiques.

C. Marques épigénétiques chez *Arabidopsis thaliana*

1. Méthylation de l'ADN

Chez les eucaryotes, l'ADN est méthylé principalement au niveau de 5-méthylcytosine (m5C). Bien que tous les organismes eucaryotes étudiés présentent de la méthylation ADN, certains organismes comme la drosophile, le nématode, la levure n'en possèdent que très peu (Lyko et al., 2000). Chez l'homme, la majorité des cytosines méthylées le sont en contexte CG alors que la méthylation est retrouvée en contexte CG, CHG et CHH (H=C, T, A) chez *A. thaliana*. Chez cette espèce, 24% des cytosines sont

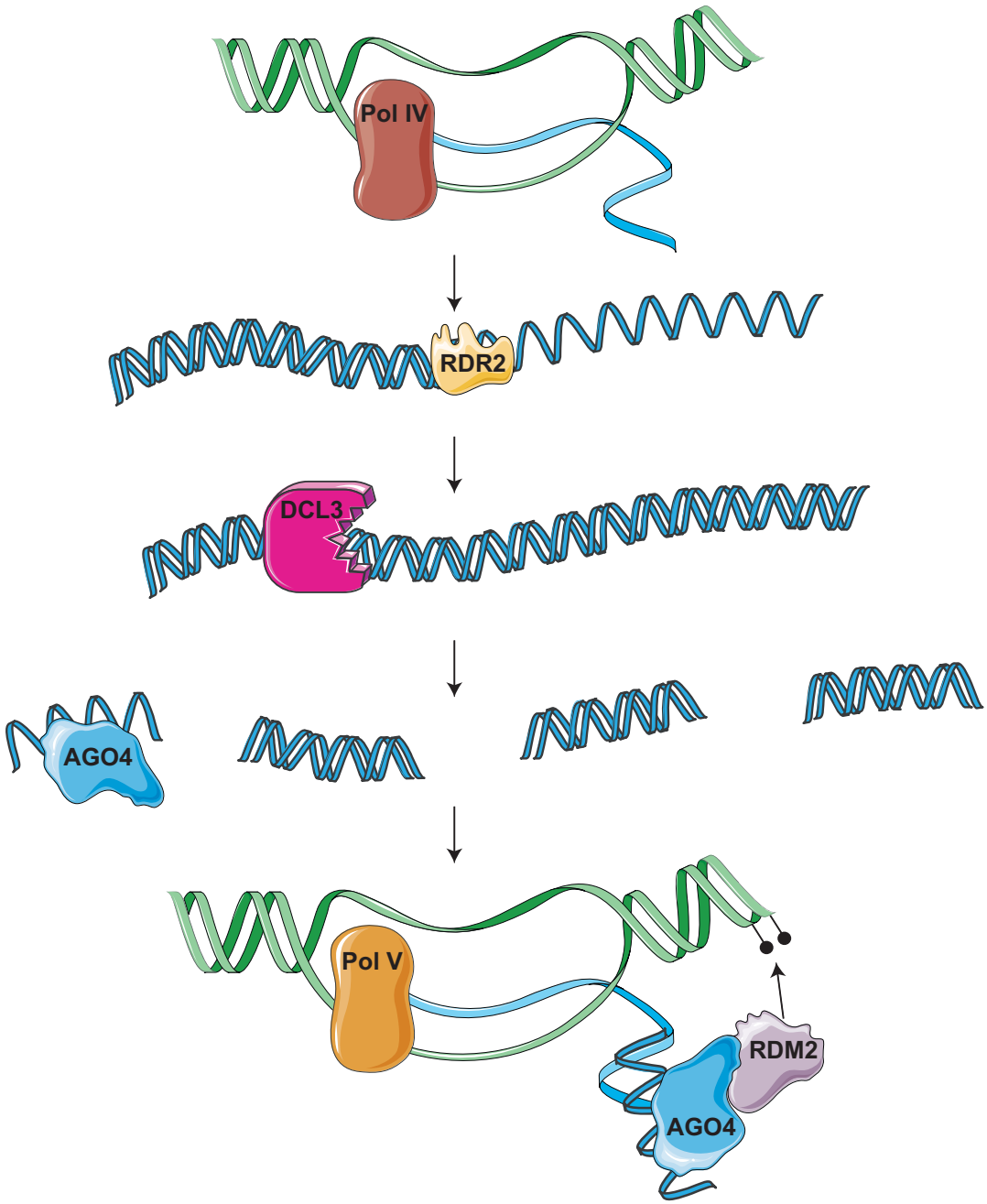


Figure 14 : La voie RNA-directed DNA Methylation (RdDM).
Chez les Végétaux, l'ARN polymérase IV produit un ARN à partir des régions d'ADN riches en séquences répétées. Cet ARN est pris en charge par RDR2 qui va former un ARN double brin à partir de l'ARN simple brin synthétisé par l'ARN polymérase IV. Cet ARN double brin va être ensuite découpé en petits ARNs de 24 nucléotides par DCL3. Un brin de ces petits ARNs va être chargé par AGO4 et va reconnaître un transcrit synthétisé par l'ARN polymérase V. L'interaction entre le petit ARN et l'ARN synthétisé par l'ARN polymérase V va permettre le recrutement de DRM2, une de novo ADN méthyltransferase qui méthyle l'ADN.

méthylées dans ce contexte CG contre 6,7% en contexte CHG et 1,7% en contexte CHH (Cokus et al., 2008). Les cytosines méthylées sont enrichies au niveau des séquences répétées des centromères et péricentromères, ainsi que dans les éléments transposables (91% des éléments transposables sont méthylés contre 20% des gènes) (Zilberman et al., 2007). Cette méthylation de l'ADN est souvent couplée à d'autres marques épigénétiques répressives portées par les histones.

Chez *A. thaliana*, la méthylation est maintenue en contexte CG par l'enzyme MET1 (METHYLTRANSFERASE 1) et est favorisée par VIM1 (VARIATION IN METHYLATION 1), VIM2 et VIM3 (Woo et al., 2007, 2008). Pour le contexte CHG, par CMT3 (CHROMOMETHYLASE 3) et pour les contextes CHH par CMT2 (CHROMOMETHYLASE 2) (Stroud et al., 2014).

Chez *A. thaliana*, la mise en place de la méthylation se fait principalement par la voie du RdDM (RNA-directed DNA Methylation) spécifique des plantes et qui fait intervenir de très nombreux partenaires protéiques telles que des ARN polymérases plantes spécifiques ainsi que des petits ARN (small ARN, sARN). Tout d'abord, l'ARN Polymérase IV (ARN Pol IV) produit des ARN simple brin à partir des séquences répétées du génome qui vont être pris en charge par RDR2 (RNA-DEPENDENT RNA POLYMERASE 2). RDR2 réplique les ARN simple brin en ARN double brin qui vont être coupés en petits ARN double brin de 24 nucléotides par DCL3 (DICER-Like 3). La protéine AGO4 va prendre en charge ces sARN de 24 nucléotides et interagir avec le transcrit généré par l'ARN Pol V permettant le recrutement de la méthyltransférase DRM2 (Domains Rearranged Methyltransferase 2) au niveau de l'ADN (Figure 14) (He et al., 2011; Law and Jacobsen, 2010; Naumann et al., 2011).

La méthylation de l'ADN est affectée la position des nucléosomes. En effet les méthyltransférases méthylent préférentiellement l'ADN enroulé autour d'un nucléosome (Chodavarapu et al., 2010). Il existe donc un lien entre organisation de la chromatine et méthylation ADN.

2. Variants d'histones

Les histones sont aussi des porteurs de l'information épigénétique notamment par leur séquence primaire en acides aminés. Chez les plantes et les mammifères, il existe des histones canoniques et des variants d'histones (Talbert and Henikoff, 2010). *Arabidopsis* possède des variants pour toutes les histones (H1, H2A, H2B et H3) sauf l'histone H4

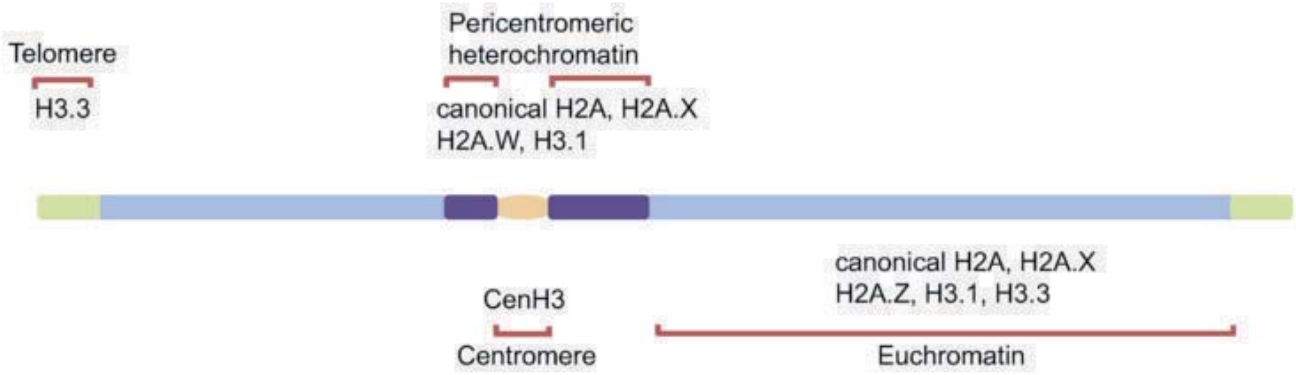


Figure 15 : Localisation des variants d'histones le long du chromosome.
Les variants d'histones présentent des caractéristiques biochimiques spécifiques ainsi que des régulations transcriptionnelles et des chaperonnes distinctes qui leurs confèrent une localisation préférentielle le long du génome (Jiang and Berger, 2016).

(Talbert and Henikoff, 2010). Les histones canoniques sont surtout transcrits en phase S ce qui va permettre leur incorporation lors de la formation de la chromatine au niveau de l'ADN nouvellement synthétisé. Les variants d'histones sont quant à eux exprimés faiblement tout le long du cycle cellulaire indépendamment de la réplication. Ils sont aussi appelés histones de remplacement (Ahmad and Henikoff, 2002).

Les variants diffèrent des histones canoniques par quelques acides aminés ou par des domaines protéiques. En modifiant la composition du nucléosome, ils influencent sa stabilité et confère une propriété plus ou moins compacte à la chromatine (Stroud et al., 2012). La distribution des variants d'histones n'est pas aléatoire dans l'euchromatine ou l'hétérochromatine et joue un rôle dans l'organisation de la chromatine (Figure 15) (Stroud et al., 2012; Yelagandula et al., 2014).

L'histone H3

Les variants (H3.3, H3.3-like et CenH3) et la forme canonique de l'histone H3 (H3.1) ont été bien caractérisés.

L'histone canonique H3.1 et le variant H3.3 ont des rôles antagonistes dans le génome. L'histone H3.1 est codée par 5 gènes (*HISTONE THREE RELATED (HTR)1* : At5g65360, *HTR2* : At1g09200, *HTR3* : At3g27360, *HTR9* : At5g10400, *HTR13* : At5g10390), et le variant d'histone H3.3 est codé par 3 gènes (*HTR4* : At4g40030, *HTR5* : At4g40040, *HTR8* : At5g10980) (Talbert et al., 2012). Les ARN synthétisés par ces gènes sont polyadénylés et pour les histones H3.1, dépourvus d'introns (Chaubet et al., 1992). Seuls quatre acides aminés en position 31/41/87/90 distinguent H3.3 et H3.1. Cette variation de la séquence primaire ainsi que leur régulation transcriptionnelle lors de la phase S, a sans doute permis leur spécialisation. La localisation des protéines H3.1-GFP et H3.3-GFP a montré que H3.1 est enrichi au niveau des chromocentres et donc au niveau de l'hétérochromatine alors que H3.3 est présent sur toute la chromatine (Ingouff et al., 2010). Des analyses d'immunoprécipitation de la chromatine (ChIP) menées à l'échelle du génome (ChIP-Seq) (Sequeira-Mendes et al., 2014; Stroud et al., 2012; Vaquero-Sedas and Vega-Palas, 2013; Wollmann et al., 2012) ont conforté ces résultats et apporté quelques précisions. H3.1 est enrichie au niveau des régions centro- et péricentromériques. Sa présence est positivement corrélée avec la méthylation de l'ADN et anti corrélée à l'expression des gènes. A l'inverse, H3.3 est enrichie dans les bras euchromatiques des chromosomes et au niveau des gènes, surtout en amont du signal de terminaison de la

transcription (TTS) (Stroud et al., 2012), mais aussi au niveau des promoteurs (Shu et al., 2014). Sa présence est positivement corrélée à l'expression des gènes et à l'occupation par l'ARN Pol II. Les nucléosomes contenant le variant H3.3 seraient moins stables que les nucléosomes portant l'histone canonique H3.1 ce qui permettrait la formation et le maintien d'une chromatine ouverte facilitant la dissociation du nucléosome lors du passage de l'ARN Pol II (Jin and Felsenfeld, 2007).

Le variant CenH3 est présent uniquement au niveau des séquences 180 pb dans les centromères (Nagaki et al., 2003; Shibata and Murata, 2004) et est indispensable à la formation des centromères et à la bonne ségrégation des chromosomes (Ingouff et al., 2010; Karimi-Ashtiyani et al., 2015; Lermontova et al., 2011).

Les variants d'histones H3.3-like (H3.14 et H3.10) sont présents dans les noyaux de cellules reproductives que sont la cellule végétative (H3.14) et dans les cellules spermatiques (H3.10) du pollen (Ingouff et al., 2007, 2010), où ils sont associés à une chromatine ouverte. Le noyau végétatif est très décompacté et ne présente pas de chromocentres alors que ceux des cellules spermatiques sont au contraire très condensés.

L'histone H2A

Il existe différents variants d'histone H2A (H2A.X, H2A.W, H2A.Z) qui, comme pour H3, possèdent des localisations génomiques et des caractéristiques spécifiques.

Le variant d'histone H2A.X est présent sur tout le génome et est très étudié aujourd'hui de par son rôle dans la signalisation et la réparation des cassures doubles brins de l'ADN. En réalité, c'est la phosphorylation de ce variant par la phosphorylase ATM (ataxia-telangiectasia mutated) qui va être un signal des dommages ADN et va permettre le recrutement de la machinerie de réparation (Yuan et al., 2010). Lorsque la réparation à été effectuée, c'est le complexe FACT (facilitates chromatin transcription) qui va remplacer le variant d'histone phosphorylé gamma H2A.X par l'histone canonique H2A (Zhou et al., 2015).

Le variant d'histone H2A.W est codé par trois gènes (AT5G02560, AT5G27670, AT5G59870). Des analyses de ChIP-seq ont montré que H2A.W est appauvri au niveau du corps des gènes et est enrichi au niveau de l'hétérochromatine ce qui a été confirmé par l'utilisation d'une fusion H2A.W-GFP qui se localise au niveau des chromocentres. *In vitro*, H2A.W favorise l'interaction des fibres de chromatine et *in vivo*, les mutants *h2a.w* montrent

une décondensation de l'hétérochromatine. A l'inverse, une expression exogène des variants H2A.W favorise la formation de chromatine condensée (Yelagandula et al., 2014). H2A.W participe donc à la formation de l'hétérochromatine.

Le variant d'histone H2A.Z est codé par trois gènes (AT2G38810, AT1G52740, AT3G54560) (Yi et al., 2006). Ce variant possède l'activité la moins bien décrite des variants H2A. H2A.Z est localisée de manière diffuse dans le noyau mais est exclue des chromocentres (Zilberman et al., 2008). Ce variant est enrichi en 5' des gènes notamment au niveau du TSS, et faiblement enrichi en 3' des gènes et est présent en quantité variable dans le corps des gènes (Coleman-Derr and Zilberman, 2012). Il a aussi été observé une corrélation entre l'enrichissement en H2A.Z et le niveau d'expression des gènes (Yelagandula et al., 2014). La présence de H2A.Z est anti-corrélée avec la méthylation de l'ADN. H2A.Z est enrichie en 5' des gènes et sur certains transposons dans le mutant *met1* mais les mutants *h2a.z* n'affectent pas la méthylation. Il semble donc que le dépôt d'H2A.Z soit sensible à la méthylation de l'ADN mais que la méthylation de l'ADN ne dépendra pas de la présence de H2A.Z (Coleman-Derr and Zilberman, 2012).

Le variant H2A.Z a un rôle répressif, sa présence est anti-corrélée avec le variant d'histone H3.3 qui est lui associé à la transcription des gènes (Stroud et al., 2012), et lorsqu'il est enrichi dans le corps des gènes, ceux ci sont peu transcrits. De plus, H2A.Z forme des nucléosomes qui enroulent l'ADN de manière plus serrée que H2A (Kumar and Wigge, 2010). Il a aussi été montré que H2A.Z est enrichie dans les corps des gènes de réponse aux stress ou aux signaux hormonaux et aussi au niveau des TSS des gènes de réponse à la chaleur (Kumar and Wigge, 2010). Lorsque les plantes sont soumises à un stress thermique, une perte d'histone H2A.Z est observée au niveau des TSS des gènes thermosensibles et cette perte est associée à une augmentation de l'expression. H2A.Z faciliterait une augmentation de transcription lorsque cela est nécessaire (Coleman-Derr and Zilberman, 2012).

Les chaperonnes d'histone

Les histones étant des protéines basiques qui interagissent fortement avec les acides nucléiques (ARN et ADN), elles ne sont jamais libres dans le cytoplasme mais doivent être prises en charge par plusieurs chaperonnes qui participent à leur stockage dans le cytoplasme, leur transport dans le noyau, leur incorporation ou encore leur éviction de la

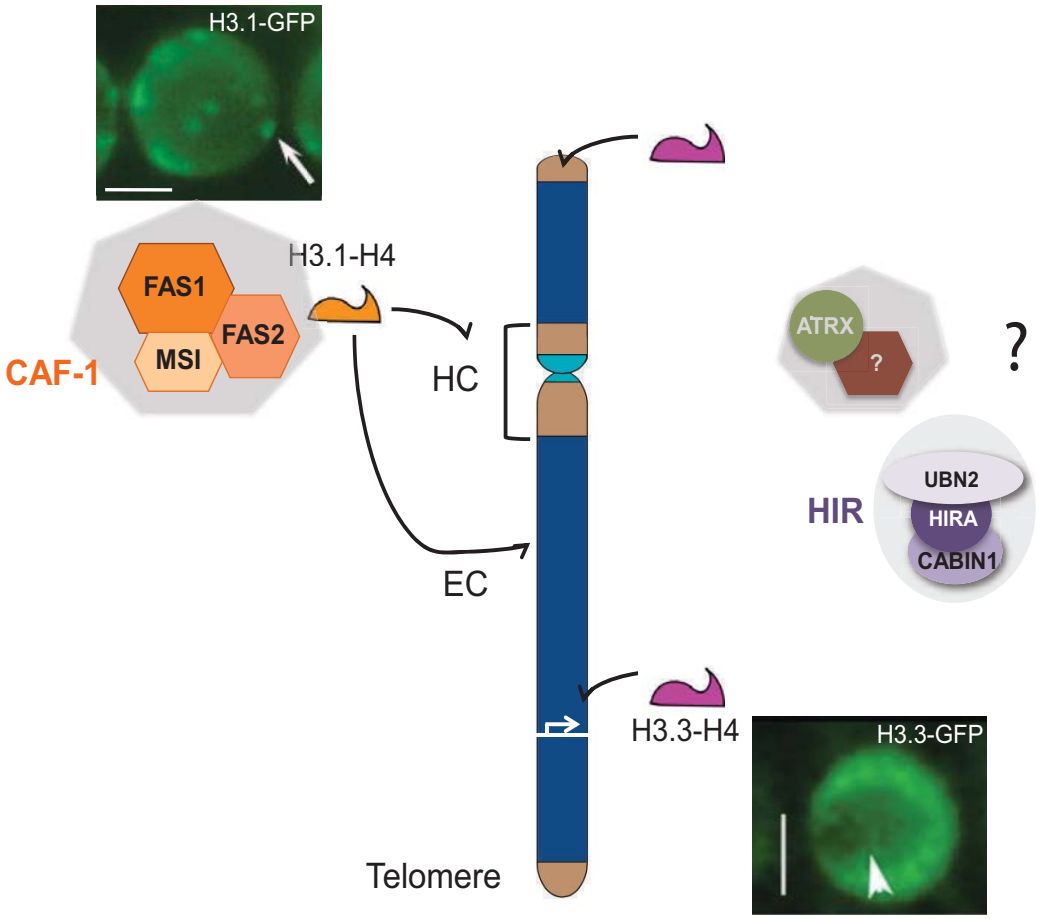


Figure 16 : Représentation schématique des mécanismes et des protéines impliquées dans le dépôt des histones H3-H4 sur l'ADN chez *Arabidopsis*.
La prise en charge des histones, leur transport et leur dépôt fait intervenir un grand nombre d'acteurs. Le complexe CAF-1 dépose l'histone canonique H3.1 lors de la réplication tout le long du chromosome. Le variant H3.3 est déposé au niveau de l'euchromatine et des télomères de façon indépendante de la réplication. Chez les mammifères, les complexes HIR et ATRX/DAXX sont responsables de ce dépôt.

chromatine (Figure 16). Je me limiterai ici à une introduction des informations sur le complexe CAF-1 et HIR disponibles au début de ma thèse.

Le complexe CAF 1 (Chromatin Assembly Factor 1) est composé de trois unités : FASCIATA1 (FAS1), FASCIATA2 (FAS2) et MULTICOPY SUPPRESSOR OF IRA1 (MSI1) (Hennig et al., 2003; Kaya et al., 2001; Ottoline Leyser and Furner, 1992). FAS1 présente un pic d'expression pendant la phase S ce qui est cohérent avec son rôle dans la formation de la chromatine pendant la réplication (Kaya et al., 2001) et est exprimé principalement dans les tissus en prolifération (Ramirez-Parra and Gutierrez, 2007a). Les mutants *fas1* et *fas2* sont viables mais présentent des défauts morphologiques au niveau de l'organisation de certains organes (trichomes, méristèmes,...) ainsi qu'une diminution du contenu en hétérochromatine du noyau (Schonrock et al., 2006) et une dé-répression modérée de la transcription des séquences répétées (Ono et al., 2006; Schonrock et al., 2006). Des expériences de ChIP réalisées avec un anticorps qui précipite l'histone H3 canonique et le variant indiquent que CAF 1 maintient la densité nucléosomale et joue un rôle dans la répression transcriptionnelle des éléments transposables après un stress thermique prolongé (Pecinka et al., 2010). Le rôle du complexe CAF 1 dans la dynamique des histones et sa potentielle spécificité pour l'histone canonique H3.1 par rapport au variant H3.3 restent à définir.

AtHIRA (At3g44530) a été identifiée comme l'une des sous-unité du complexe HIR (Ingouff et al., 2010; Phelps-Durr et al., 2005). Deux mutants par insertion T-DNA de ce gène ont été décrits comme létaux (Phelps-Durr et al., 2005) et un autre comme ne montrant pas de phénotype (Ingouff et al., 2010). Chez l'homme, le complexe HIR comprend trois sous unités : HIRA, CABIN et UBINUCLEIN et interagit directement avec l'ADN (Ray-Gallet et al., 2011). Le rôle de HIRA dans l'organisation de la chromatine chez les plantes et l'identification des potentiels composants du complexe HIR restaient donc à découvrir au début de mon travail de thèse.

Bien que les connaissances sur la dynamique des histones et les différents complexes impliqués se soient multipliées ces dernières années, beaucoup reste à caractériser chez les plantes surtout concernant les rôles de protéines chaperonnes d'histones dans la dynamique des variants d'histones et leurs rôles dans la stabilité et la régulation transcriptionnelle du génome.

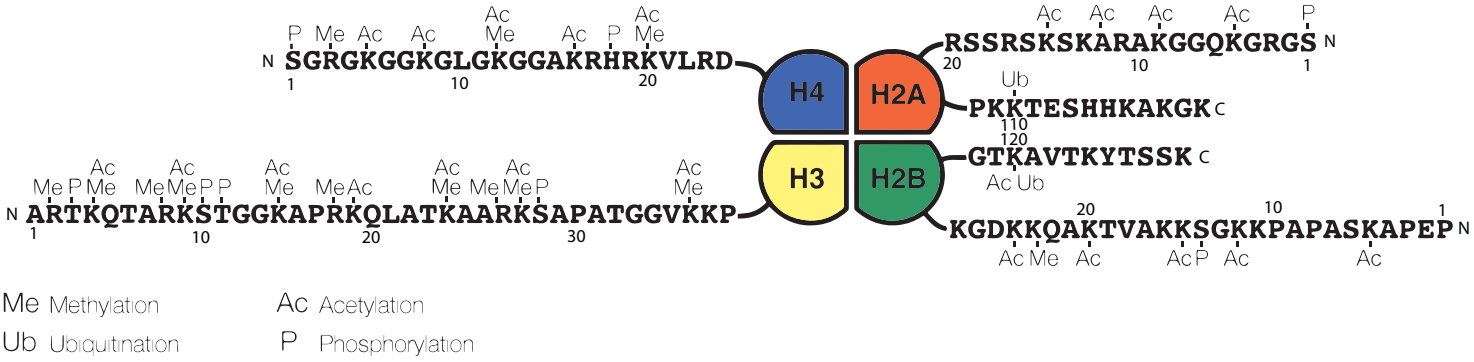


Figure 17 : Représentation schématique du « code histone ».
Les queues N-terminales des histones peuvent subir des modifications post-traductionnelles (PTMs) et certaines combinaisons génèrent un "code histone" qui défini un état chromatinien. La nomenclature des PTMs correspond au nom de l'histone modifiée, l'acide aminé qui porte cette modification ainsi que sa position dans la séquence protéique et enfin le type de modification. Exemple : H3K4me2 : di-méthylation de la lysine (K) 4 de l'histone H3.

3. Les modifications post-traductionnelles des histones

Les histones peuvent subir de nombreuses modifications au niveau de leurs extrémités (« queues des histones ») et leur domaine globulaire telles que des acétylations, des méthylations, des ubiquitinations, des phosphorylations et beaucoup d'autres (Kouzarides, 2007). Il existe un réel code d'histones qui n'est pas encore totalement décodé (Figure 17). Ces modifications peuvent influencer la stabilité du nucléosome et donc participer à la régulation de la compaction de la chromatine ce qui aura des conséquences sur le niveau de transcription. De plus, ces modifications post-traductionnelles peuvent être lues par des protéines lectrices (« readers ») qui peuvent entraîner le dépôt d'autres marques épigénétiques. Notons ici que les marques d'histones peuvent être conservées entre les organismes ou exister dans un organisme et pas dans un autre, ou encore présenter des caractéristiques différentes entre les organismes. L'euchromatine est enrichie en marques d'histones dites actives ou permissives pour la transcription telles que H3K4me3 et H3K9ac, alors que l'hétérochromatine est enrichie en marques répressives telles que H3K9me2, H3K27me1... Ces modifications post-traductionnelles (Post-Translational Marks ; PTMs) sont réversibles et leurs dépôts et retraits sont effectués par des enzymes spécifiques.

Certaines PTMs influencent directement la biochimie des nucléosomes comme c'est le cas des acétylations qui inhibent les charges positives des lysines de la queue N-terminale des histones diminuant leur interaction avec l'ADN et forment une chromatine plus ouverte dite « permissive » et favorable pour la transcription (Bannister and Kouzarides, 2011). Inversement, certaines PTMs dites « répressives » stabilisent l'interaction histone-ADN et répriment la transcription.

- *Marques répressives*

La marque H3K9me2 est enrichie au niveau des régions centro- et péricentromériques ainsi que sur les éléments transposables présents au sein de l'euchromatine (Roudier et al., 2011b). Les histones méthyltransférases SUPPRESSOR OF VARIEGATION (SU(VAR)3-9) HOMOLOGS (SUVH) 4/5/6 sont responsables du dépôt de la marque H3K9me2 (Ebbs et al., 2005; Ebbs and Bender, 2006; Jackson et al., 2002). La méthylation H3K9me2 est déposée, puis amplifiée en boucle. En effet la marque H3K9me2 est reconnue par l'ADN méthyltransférase CHROMO METHYLTRANSFERASE 3 (CMT3) qui va apposer la méthylation de l'ADN en contexte CHG sur les domaines de chromatine

enrichis en H3K9me2. En retour, les protéines SUVH reconnaissent la méthylation CHG et déposent H3K9me2 (Du et al., 2012).

Un autre exemple de marques répressives pour la transcription est la marque H3K27me1 qui est associée à l'hétérochromatine péricentromérique (Jacob and Feng, 2009; Mathieu et al., 2005; Roudier et al., 2011b). Cette marque est déposée par les méthyltransférases ARABIDOPSIS TRITHORAX-RELATED (ATXR) 5 et 6 (Jacob et al., 2010, 2014; Jacob and Feng, 2009). Le double mutant *atxr5 atxr6* montre une décondensation des chromocentres et une levée de répression sur les éléments transposables (Jacob and Feng, 2009). ATXR5 et 6 sont importants afin d'éviter la re-réplication des séquences répétées (Jacob et al., 2010). *In vitro*, les méthyltransférases ATXR 5 et 6 vont préférentiellement méthyler l'histone canonique H3.1 plutôt que le variant H3.3. Cette découverte démontre l'existence d'une interaction entre les marques épigénétiques telles que la méthylation des histones et les variants d'histones (Jacob et al., 2014).

- *Marques permissives*

Il existe aussi des marques permissives pour la transcription telles que H3K4me3. Cette marque est associée aux gènes fortement exprimés (Zhang et al., 2009). H3K4me3 est enrichie au niveau du promoteur et de la partie 5' de la région transcrite des gènes. Cette marque est déposée majoritairement par la protéine SET DOMAIN GROUP 2 (SDG2) (Guo et al., 2010). La mutation *sdg2* entraîne une perte drastique de la marque H3K4me3 *in vivo* et une dérégulation massive de nombreux gènes (Guo et al., 2010). En plus de SDG2, le complexe ATX1 (ARABIDOPSIS HOMOLOGY OF TRITHORAX 1) / COMPASS et la protéine ATX7 semblent aussi intervenir dans la méthylation d'H3K4me3 mais avec une action limitée à certains gènes (Tamada et al., 2009).

H3K36me3 est un autre exemple pour une marque associée à la transcription, elle est déposée par SDG8 (Xu et al., 2008; Zhao et al., 2005). Un autre membre de la famille, SDG4, qui est exprimé de manière spécifique dans les fleurs, est impliqué dans le dépôt des marques H3K4me3 et H3K36me3 (Cartagena et al., 2008).

Certaines marques d'histones sont impliquées dans plusieurs processus cellulaires comme la marque H3K56ac. Cette marque est associée à la transcription des gènes, mais aussi à la réplication (Lee et al., 2010). Lorsqu'elle est associée à la transcription, H3K56ac

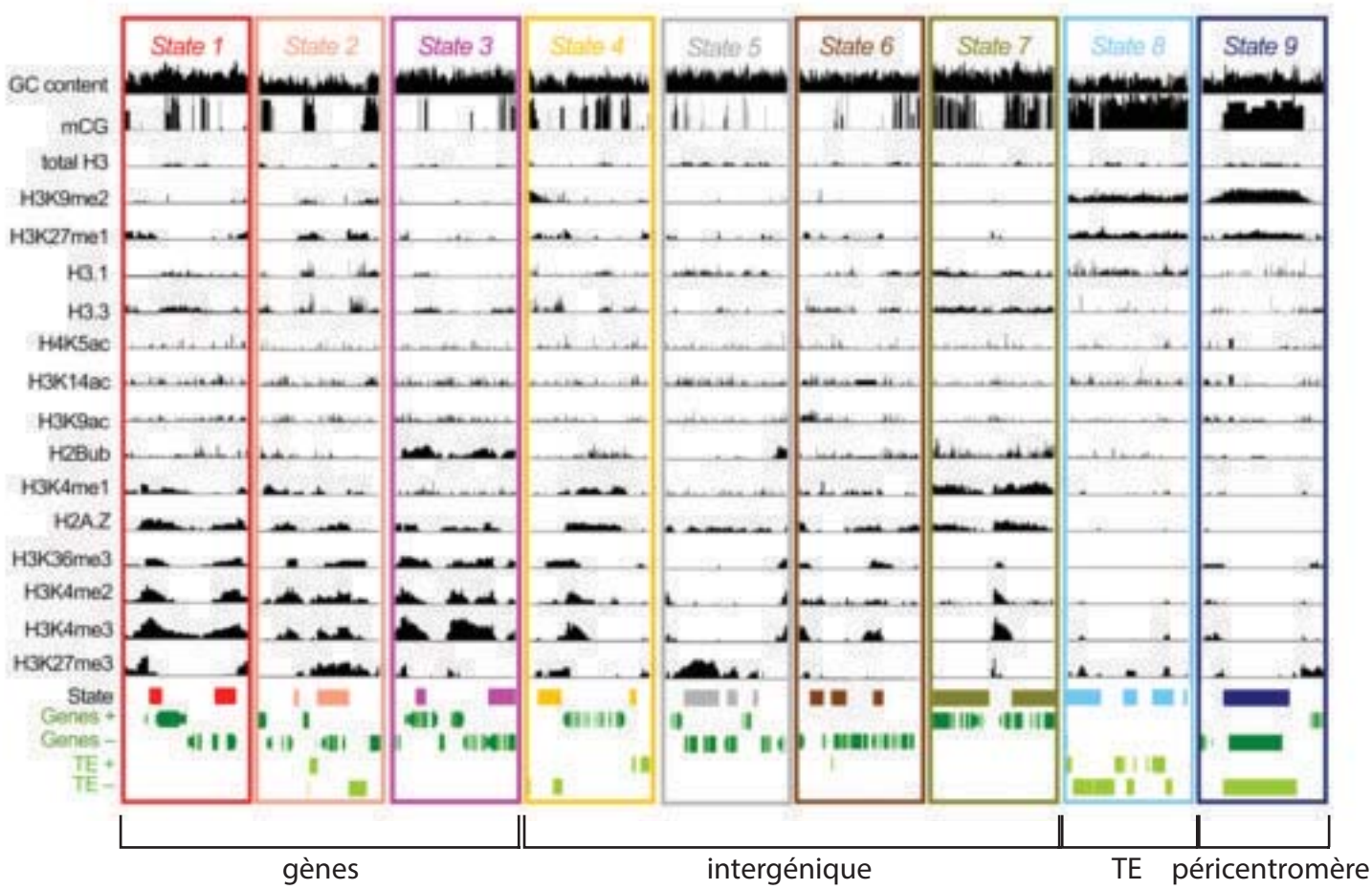


Figure 18 : Représentation de *loci* génomiques avec différentes marques épigénétiques définissant les états chromatiniens.
L'état chromatinien 1 (CS1) est riche en H3K4me2, H3K4me3, H3ac, H3K36me3 et il est associé à la région transcrite des gènes et au TSS. C'est un état de faible densité nucléosomale enrichi en histones H3.3 et H2A.Z. Les mêmes marques actives sont trouvées au niveau du CS2, mais ce état chromatinien est aussi enrichi en marques répressives H3K27me3 et présente moins de H3K36me3, H2Bub et H3ac. CS3 est très proche de CS1 (haut niveau de H3K4me1, H2Bub, H3K36me3 et H3K4me2/3) mais les marques épigénétiques ne sont pas enrichies au niveau des promoteurs et de la région 5' des gènes. CS4 ressemble à CS2 avec un enrichissement en H3.3, H2A.Z et H3K27me3 mais avec moins de marques actives. CS4 est présent au niveau des régions non codantes. CS5 présente peu de marques, mais il est très enrichi en H3K27me3 et faiblement en H2A.Z et H3.1. Il correspond majoritairement à des régions intergéniques. CS6 correspond aussi aux régions intergéniques avec un enrichissement en H2A.Z, une forte densité nucléosomale au niveau du corps des gènes. CS7 correspond aussi aux régions intergéniques et présente les marques actives H3K4me1, H2Bub et H3K36me3. CS8 et CS9 sont des états hétérochromatiniens qui se distinguent par leur contenu en GC. CS9 est riche en GC par rapport à CS8. CS8 est associé au TE et aux zones intergéniques alors que CS9 est associé aux péricentromères. D'après Sequeira-Mendes and Gutierrez, 2015.

est enrichie au niveau des promoteurs et en 5' de gènes (Roudier et al., 2011b; Tanurdzic et al., 2008) alors que lorsqu'elle est associée à la réplication elle est positionnée au niveau des régions intergéniques. Cette marque est souvent corrélée à la marque H3K4me3 au niveau de l'euchromatine.

- *Domaines de chromatine*

L'analyse de l'enrichissement en différentes combinaisons de marques épigénétiques dans le génome d'*A. thaliana* a permis de définir des états chromatiniens (CS: Chromatin State) :

(Roudier et al., 2011b) ont d'abord identifié quatre états chromatiniens :

CS1 qui est associé avec H3K4me2, H3K4me3, H3K36me3, H3K9me3, H2Bub, et H3K56ac et CS2 qui est associé à H3K27me2 et H3K27me3. Ces deux états de la chromatine sont majoritairement associés à des gènes, CS1 étant associé à un état transcriptionnellement actif et CS2 à des gènes réprimés ou faiblement exprimés. CS3 est l'état transcriptionnel associé aux éléments transposables et aux centromères, il est enrichi en marques H3K9me2, H4K20me1, H3K27me1, H3K27me2 et à la méthylation de l'ADN. CS4 n'est pas enrichi particulièrement en marques épigénétiques et correspond aux zones intergéniques et aux gènes faiblement exprimés (Roudier et al., 2011b). Ces quatre états ont été affinés par la suite en prenant en compte d'autres marques épigénétiques, la distribution génomique d'H3.1 et H3.3 ainsi que la séquence ADN (Figure 18) (Sequeira-Mendes et al., 2014).

Ces différents états montrent la complexité des mécanismes qui interviennent dans la régulation de la chromatine.

IV. La dynamique de la chromatine chez *Arabidopsis thaliana*

La chromatine est dynamique, elle n'adopte pas une organisation statique et change en réponse aux *stimuli* extérieurs ou en fonction du stade de développement.

Lors du développement, il existe plusieurs stades pendant lesquels un changement important de l'organisation de la chromatine est observé.

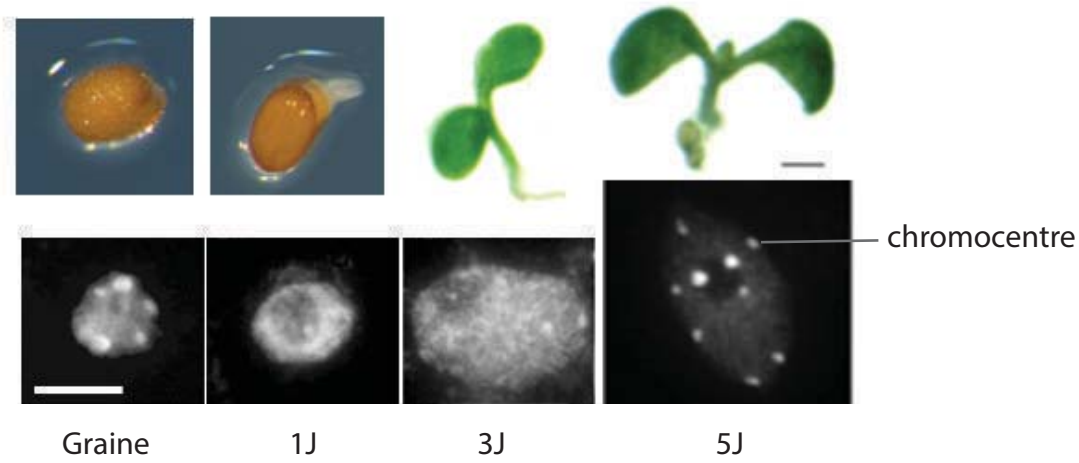


Figure 19 : Organisation de la chromatine au sein du noyau pendant le développement précoce de la plante.
Photographies représentatives des stades de développement d'Arabidopsis thaliana de la graine jusqu'à 5 jours après germination ; les colorations au DAPI des noyaux associées à chaque stade sont présentées. Les chromocentres se décondensent pendant la germination et se re-condensent progressivement pendant les premier jours après la germination jusqu'à atteindre une forme mature à 5 jours après germination. D'après Benoit et al., 2013 et van Zanten et al., 2011.

Par exemple, lors de la transition florale (le moment où la plante passe du stade végétatif au stade reproducteur) une décondensation des chromocentres associée à une décompaction des séquences répétées est observée dans les feuilles 4 jours avant l'émergence de la hampe florale (Tessadori et al., 2007b). Cette décompaction concerne majoritairement l'ADNr 5S alors que les séquences centromériques 180 pb et l'ADNr 45S restent condensées. Il semble donc que les différentes séquences répétées du génome sont affectées de manières différentes par la transition florale. Cette décondensation n'implique pas de changement majeur au niveau de la méthylation de l'ADN car les séquences répétées restent fortement méthylées. Les chromocentres se reforment une fois l'élongation de la hampe florale achevée.

D'autres changements majeurs de l'organisation chromatinienne ont été observés dans les cotylédons (Douet et al., 2008; Mathieu et al., 2003a; van Zanten et al., 2011). Les cotylédons sont les premiers tissus aériens formés par la plante. Ils sont formés pendant l'embryogénèse et se développent après la germination (Chandler, 2008). Il a été montré que les cotylédons embryonnaires (à l'intérieur de la graine) présentent une structure chromatinienne similaire à celle d'un cotylédon de plante adulte (van Zanten et al., 2011). Pendant la maturation de la graine et son assèchement, la chromatine se condense de plus en plus, avec les séquences péricentromériques qui se regroupent dans les chromocentres (van Zanten et al., 2011). Cette très forte compaction de la chromatine dans la graine est associée à une très faible expression des gènes. Pendant la germination, l'hétérochromatine se décondense avec une dispersion des séquences répétées dans le noyau. De très petits chromocentres appelés "pre-chromocentres" sont observés dans les noyaux de cotylédons juste après la germination. Ces structures précurseurs des chromocentres sont déjà enrichis en méthylation de l'ADN et en marque répressive H3K9me2 (Douet et al., 2008; Mathieu et al., 2003a). Dans les premiers jours après la germination, les cotylédons se développent et la chromatine se recondense dans les noyaux : les chromocentres grandissent jusqu'à présenter un aspect similaire à ceux d'une feuille adulte environ à 4 jours après germination (Figure 19). La croissance du cotylédon entre 2 et 4 jours après germination correspond à la transition hétérotrophe/autotrophe et implique très probablement des changements du profil d'expression de nombreux gènes. Ces changements pourraient être accompagnés d'une altération de la structure en rosette des chromocentres. Il reste toutefois de nombreuses inconnues et le rôle joué par les chromocentres dans cette reprogrammation transcriptionnelle est encore largement inconnue.

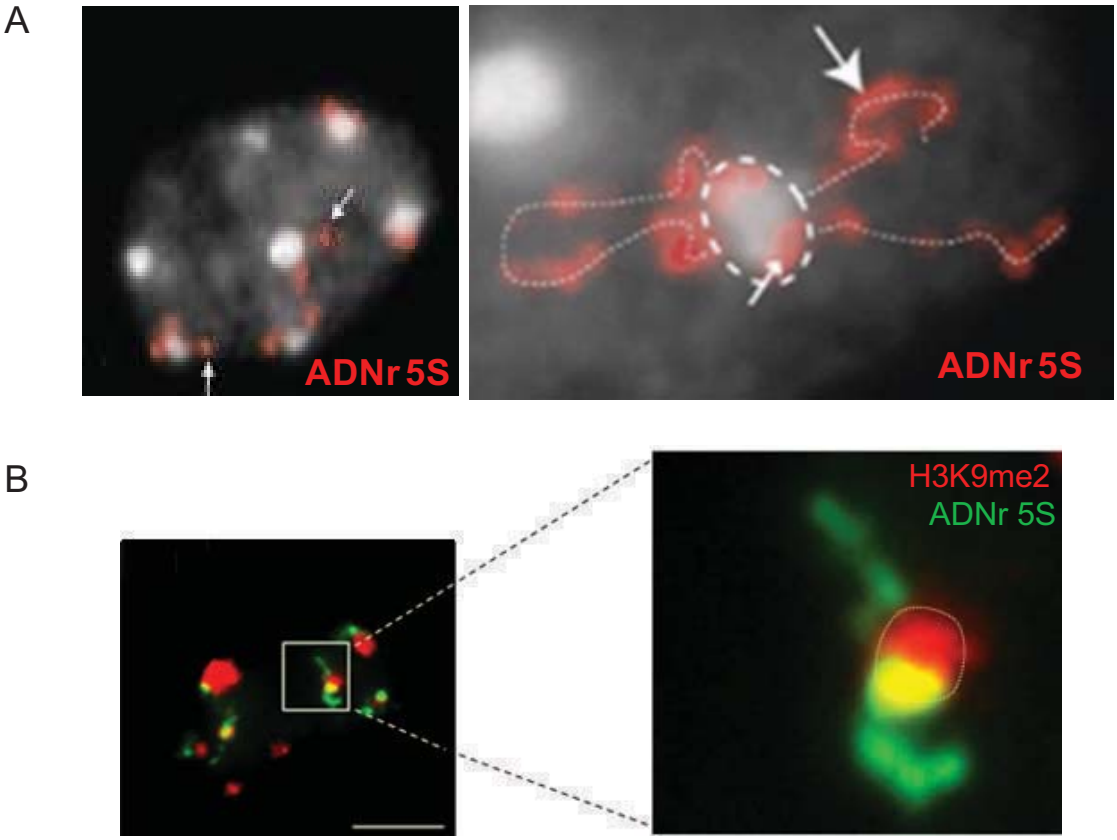


Figure 20 : Organisation des gènes d'ARNr 5S dans le noyau.
A) Une partie des gènes d'ARNr 5S est présente au niveau des chromocentres et une autre partie des gènes est localisée sur les boucles euchromatiques.
B) Les gènes d'ARNr 5S présents dans le chromocentre sont enrichis en marque H3K9me2 contrairement à ceux présents sur les boucles.

V. Organisation chromatinienne et ADNr chez *A. thaliana*

A. Organisation des ADNr dans un noyau en interphase

Des analyses de FISH et d'immuno-FISH ont montré que, dans un noyau de feuille en interphase, les gènes d'ARNr 5S sont localisés d'une part à l'intérieur des chromocentres et associées à la marque répressive H3K9me2 et d'autre part en "boucles" plus externes non associés à H3K9me2 (Figure 20) (Mathieu et al., 2003a). Dans les mutants de la voie RdDM, on observe une augmentation de la transcription des ARNr 5S minoritaires ainsi qu'une augmentation de la taille des boucles d'ADNr 5S. Ces observations suggèrent que les gènes d'ARNr 5S situés dans les boucles d'ADNr seraient transcriptionnellement actifs. Toutefois, dans le mutant *met1-1* on observe la décondensation de l'hétérochromatine mais pas d'augmentation de la transcription des ARNr 5S mineurs alors qu'inversement le mutant *ago4-1* (qui joue un rôle dans le RdDM) augmente la transcription des ARNr 5S mineurs sans changer le contenu en hétérochromatine (Vaillant et al., 2007). La décompaction de l'hétérochromatine et la transcription des gènes d'ARNr 5S ne sont donc pas nécessairement toujours corrélées.

B. Organisation des gènes d'ARNr pendant le développement

Les gènes d'ARNr 5S subissent une large réorganisation au sein du noyau entre 2 et 5 jours (J) après la germination (Douet et al., 2008). À 2J, ils co-localisent avec les pré-chromocentres. A 3J, les gènes d'ARNr 5S se décondensent et forment des boucles plus externes par rapport au chromocentre (Douet et al., 2008; Mathieu et al., 2003a). Cette décondensation est suivie par une re-condensation progressive des gènes d'ARNr 5S. La décondensation est accompagnée d'une déméthylation des ADNr 5S par la déméthylase ROS1 (Repressor of silencing 1) alors que la condensation nécessite l'activité de la polymérase IV (Douet et al., 2008).

La transcription de l'ARNr 5S étant contrôlée par le facteur de transcription TFIIIA, des études de corrélation entre le niveau d'expression de l'ARNr 5S et de TFIIIA ont été réalisées (Layat et al., 2012). Entre 1J et 3J, TFIIIA est très faiblement exprimé et les protéines TFIIIA produites sont dégradées. En accord avec l'absence de TFIIIA, la quantité d'ARNr 5S diminue durant les premiers jours après la germination. A 4J, les chromocentres sont formés, la synthèse de la protéine TFIIIA reprend et la synthèse d'ARNr 5S augmente. Ceci suggère que la réorganisation des gènes d'ARNr 5S permet d'assurer l'expression des

ARNr 5S et en même temps la mise sous silence sélective des copies minoritaires des ARNr 5S.

Pour les gènes d'ARNr 45S, le gène du type VAR1 qui représente 50% des séquences ADN est exprimé dans les graines puis mis sous silence lors du développement de la plante pendant que les autres variants 45S sont exprimés (Earley et al. 2006, 2010; Pontvianne et al. 2010). La répression transcriptionnelle de VAR1 est levée si la méthylation de l'ADN est diminuée après application de la 5-aza-2′-deoxycytosine (5-AZA) un analogue de la cytosine, ou en présence de mutations pour certains modificateurs de la chromatine comme l'HISTONE DEACETYLASE 6 (HDA6) (Pontvianne et al., 2013). Les gènes d'ARNr 45S semblent donc également être régulés lors du développement par des modifications épigénétiques.

VI. Régulation épigénétique de l'ADNr 5S chez *A. thaliana*

A. La méthylation de l'ADN

Il a été observé par des "analyses bisulfite" que les gènes d'ADNr 5S sont fortement méthylés (79% des cytosines présentes au niveau des gènes d'ARNr 5S sont méthylés contre 8% dans le génome) et ce quelque soit le contexte de méthylation (CG, CHG, CHH) (Mathieu et al., 2002a).

Ceci dit, l'utilisation *in vivo* de 5-AZA ou la méthylation *in vitro* sur des extraits de tabac par une méthylase CG spécifique n'entraînent aucun changement de la transcription des gènes d'ARNr 5S (Mathieu et al., 2002b). La méthylation ne semble donc pas être suffisante pour réprimer la transcription des gènes d'ARNr 5S. Par contre, l'altération de la méthylation CG dans le mutant *met1* ou le remodelage chromatinien dans le mutant *ddm1* conduisent à une augmentation du nombre d'ARNr 5S minoritaires ainsi qu'une transcription des ARN 5S-210 (Mathieu et al., 2003a; Vaillant et al., 2007). Il semble donc que se soit plutôt l'état chromatinien qui soit responsable de la régulation de la transcription des ARNr 5S.

L'analyse des séquences ADN contenues dans les YACs répartis le long du grand *locus* du chromosome 5 a permis de montrer que les séquences les plus centromériques sont plus riches en méthylation asymétrique (CHH) et possèdent un rapport AT/GC plus élevé que les séquences euchromatiques (Mathieu et al., 2002a; Vaillant et al., 2008). L'une

des hypothèses pour expliquer cet enrichissement serait le mécanisme de désamination des cytosines méthylées en thymine. Ce mécanisme conduirait à une diminution au cours de l'évolution de la proportion de sites CG et CNG par rapport aux sites CNN et à l'augmentation de la fréquence des résidus A+T dans les régions méthylées des génomes.

B. La voie RdDM

Il a été montré dans l'équipe que dans des mutants de la sous unité commune aux deux ARN polymérases plantes-spécifiques et impliquées dans la voie RdDM (ARN Pol IV et Pol V), les gènes d'ARNr 5S ne se réorganisent pas au sein du chromocentre pendant le développement post-germinatif (Douet et al., 2008). D'autres études réalisées à des stades plus tardifs dans les feuilles montrent également que les *loci* d'ADNr 5S sont décondensés dans certains mutants de la voie RdDM (Douet et al., 2009; Onodera et al., 2005; Pontes et al., 2009). De plus, les mutations dans des sous unités de l'ARN Pol IV et Pol V entraînent une augmentation du transcrit de 5S-210 bases, une caractéristique de la dé-répression des gènes d'ARNr 5S (Douet et al., 2009). La dé-répression du transcrit de 210 pb dans les mutants Pol IV et V est associée à une diminution de la production des siARN 5S de 24 nucléotides (Blevins et al., 2009; Onodera et al., 2005). Les protéines AGO4 fixent les siARNs issus des ARNr 45S et 5S (Pontes et al., 2006) et les plantes mutées pour la protéine AGO4 présentent une diminution des siARNs sur les gènes d'ARNr 5S, une perte de la méthylation CHG/CHH et de marques épigénétiques répressives H3K9me2 (Herr et al., 2005; Onodera et al., 2005; Pontes et al., 2006). Dans les plantes mutantes pour Pol IV, la méthylation est également diminuée sur les gènes d'ARNr 5S (Onodera et al., 2005).

L'ARN Pol V possède en plus une action uniquement sur les gènes d'ARNr 5S du *locus* du chromosome 4, indépendante de la voie RdDM. La dé-répression observée dans les mutants Pol V est associée à une décondensation des gènes d'ARNr 5S du chromosome 4 en FISH (Douet et al., 2009). Ces résultats suggèrent également l'existence d'une régulation *locus* spécifique des gènes d'ARNr 5S.

En résumé, les gènes d'ARNr 5S sont des séquences organisées en répétition en tandem dans les régions péricentromériques des chromosomes. Bien que localisés dans ces régions péricentromériques qui sont des régions transcriptionnellement inactives, les gènes d'ARNr 5S sont fortement transcrits par l'ARN polymérase III. Dans les noyaux en interphase une partie de gènes d'ARNr 5S est localisée au sein du chromocentre et une autre partie des séquences sur les boucles euchromatiques. Il est admis que le génome encode plus de

gènes d'ARNr 5S que nécessaire afin d'assurer un niveau de transcription des ARNr 5S suffisant pour la cellule. Certains gènes d'ARNr 5S ne sont exprimés qu'à certains stades de développement et mis sous silence par des mécanismes épigénétiques à d'autres moments du développement. De plus, la transcription des gènes d'ARNr est contrôlée par l'activité de facteur de transcription TFIIIA.

Objectifs

L'ARNr 5S est un composant du ribosome et est donc un des acteurs indispensables à la vie cellulaire. Bien qu'ayant un rôle central dans la cellule, peu de données sont disponibles sur la régulation des gènes d'ARNr 5S. Cette absence de connaissance est liée au fait que les gènes d'ADNr 5S soient organisés en répétitions en tandem et soient positionnés au niveau péricentromérique, c'est-à-dire dans une région riche en séquences répétées ce qui rend leur étude difficile. Ainsi, malgré le séquençage et l'assemblage du génome d'*A. thaliana* depuis près de 16 ans (AGI, 2000), les séquences couvrant les régions centromériques et péricentromériques, notamment celles couvrant l'ARN 5S, ne sont pas assemblées. Les gènes d'ARNr 5S sont intrigants puisque malgré leur localisation dans une région plutôt silencieuse en terme de transcription, l'ARNr 5S fait partie des ARNs les plus abondants de la cellule.

De plus, seules 290 copies d'ARN 5S sur la pseudomolécule ont été recensées ce qui n'est pas cohérent avec les estimations du nombre de copies (Campell et al., 1992). La pseudomolécule est donc incomplète et nous avons donc décidé d'étudier le nombre de copies d'ADNr 5S chez *Arabidopsis thaliana*.

Mon travail de thèse s'est déroulé dans un laboratoire ayant largement participé à la compréhension de la structure et de la fonction des gènes d'ARN 5S. Plus précisément, ma thèse a eu pour objectif de mieux comprendre l'organisation et la régulation des gènes d'ARNr 5S. En effet, il a été proposé qu'une régulation épigénétique soit responsable du choix des gènes transcriptionnellement actifs, mais l'absence de séquence de référence couvrant les clusters d'ADNr 5S rend difficile les analyses et les interprétations (S. Tourmente, données non publiées).

Des travaux précédents du laboratoire ont , en effet, mis en évidence l'existence d'une séquence riche en thymine (T-stretch) en aval de la séquence transcrite de 120 bases qui permet de distinguer les trois principaux clusters de gènes d'ARNr 5S. Ces études ont été réalisées par séquençage Sanger à partir de peu de séquences. Il en est de même pour l'analyse des polymorphismes des transcrits d'ARNr 5S qui reste très préliminaire.

Un **premier objectif** de ma thèse consistait donc à valider les résultats obtenus par séquençage Sanger grâce à des analyses 'whole genome' et de caractériser de façon plus exhaustive les polymorphismes des gènes d'ARNr 5S au niveau de la séquence du T-stretch et le long de la séquence transcrite. Afin d'atteindre cet objectif, je me suis basée sur des données de séquençage génomiques disponibles puis j'ai généré mes propres données de séquençage avant de les analyser grâce à des outils bioinformatiques. Ce travail a été réalisé en collaboration avec un doctorant (Axel Poulet) et deux stagiaires de Master 2 (Tristan Dubos et Damien Lauber).

Dans un **deuxième objectif**, j'ai appliqué ces outils sur des données de ChIP-seq disponibles afin d'étudier l'enrichissement de l'ADNr 5S en différentes marques épigénétiques de manière globale et de manière plus spécifique pour chaque cluster d'ADNr 5S. Ce travail permettra de mieux comprendre le statut épigénétique et ses possibles conséquences sur la régulation transcriptionnelle des gènes d'ARNr 5S.

Enfin, dans un **troisième objectif**, je me suis intéressée aux possibles variations du nombre de copies d'ADNr 5S, leurs polymorphismes ainsi que leurs localisations génomiques dans les populations d'*A. thaliana* séquencées par le consortium des « 1001 génomes » (http://1001genomes.org/; Alonso-Blanco et al., 2016). Le but était d'augmenter nos connaissances concernant la dynamique et la variabilité des gènes d'ARNr 5S au sein du genre *Arabidopsis* ainsi que de mieux comprendre le possible impact de l'organisation chromatinienne sur cette variabilité et sur la régulation transcriptionnelle.

Résultats

Organisation, variabilité et contrôle épigénétique des gènes d'ARNr 5S chez *Arabidopsis thaliana*

L'ARNr 5S fait partie de la grande sous unité du ribosome et participe à la synthèse protéique. Chez *Arabidopsis thaliana*, il est transcrit à partir de gènes d'ARNr 5S composés d'une séquence transcrite de 120pb et d'une région intergénique de 380pb contenant une séquence riche en thymine appelée T-stretch. Ces gènes sont organisés en répétitions en tandem en trois *loci* majeurs dans les zones péricentromériques des chromosomes 3, 4 et 5.

Dans ce projet, nous avons généré et utilisé des données NGS, afin de mieux caractériser l'organisation chromatinienne et la régulation épigénétique et transcriptionnelle de ces séquences.

Nous avons déterminé la présence de plus de 2000 copies de gènes d'ARN 5S (soit deux fois plus que ce qui était jusqu'à présent admis), confirmé la présence d'un T-stretch majoritaire pour chaque *locus* et l'existence de polymorphismes sur ce T-stretch ainsi que des polymorphismes *locus*-spécifiques au sein de la séquence transcrite. Concernant les marques épigénétiques nous avons constaté un enrichissement en marques répressives H3K9me2 et H3K27me1, en histone canonique H3.1 et en variant d'histone H2A.W associés à l'hétérochromatine. Il existe aussi des différences d'enrichissement entre les trois *loci* 5S.

L'analyse de différents écotypes et de mutants impliqués dans le pathway RdDM a démontré une variabilité dans le nombre de gènes d'ARNr 5S et l'apparition ou la disparition des *loci* 5S. Nous nous sommes alors intéressés à leur origine. En utilisant des sondes LNA (dont la mise au point du protocole a fait l'objet de l'écriture d'un article cf **Annexe II**) et des amorces qPCR spécifiques d'un T-stretch, nous avons pu montrer que le *locus* ayant subit des changements (duplication ou disparition) dans les mutants et l'écotype étudiés est celui du chromosome 5.

L'identification des mutants du pathway RdDM qui montrent des altérations au niveau de nombre de *loc*i d'ADNr 5s a été réalisée par Cécilia Oliver et les scripts bioinformatiques utilisés dans l'article ont été réalisés par Tristan Dubos et Damien Lauber.

L'analyse des 1001 génomes d'*Arabidopsis* ainsi que l'exploitation des MAGIC lines a été réalisée par Fernando Rabanal (GMI, Vienna).

Copy number and epigenetic variations at 5S rRNA genes reveal genome dynamics in the Arabidopsis species

Short title:

Dynamics at 5S rRNA genes

Lauriane SIMON [1], Fernando A. RABANAL [2], Cecilia OLIVER [3], Tristan DUSBOS [1,4], Damien LAUBER [1], Axel POULET [1,5], Christophe TATOUT[1] and Aline V. PROBST[1, #]

[1] UMR CNRS 6293 - INSERM U1103 – GReD, Clermont Université, 63178 Aubière

[2] Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria

[3] Present Address: UPR1142, Institut de génétique humaine (IGH), 34000 Montpellier

[4] Present Address: CHRU Brabois, 54000 Nancy

[5] Present Address: Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta GA 30322

Corresponding authors: aline.probst@univ-bpclermont.fr

Abstract

Background:

5S rRNA is a critical component of the ribosome and its synthesis is tightly regulated to suit the cellular requirements. 5S rRNA is encoded by gene arrays organized in tandem repeats situated in the pericentromeric regions of chromosomes 3, 4 and 5 in the Arabidopsis Col-0 genome. Full genome assembly remains challenging in these highly repeated regions and impedes further investigation of their organization, evolutionary dynamics and epigenetic regulation.

Results:

Here we provide in depth information on 5S rRNA genes in *Arabidopsis thaliana*. Specific DNA signatures were recorded in next-generation sequencing datasets to define chromosome-specific polymorphisms and reference sequences. 5S rRNA genes are enriched in repressive marks whereas the loci on chromosome 4 and 5 also display peculiar histone modifications and variants characteristic of transcriptional activity. We report that 5S rDNA loci are highly dynamic with alterations in chromosomal positions through translocations and important copy number variations between different ecotypes in the 1001 genome population. Using the L*er* ecotype, in which reorganization events were recorded, and specific mutant backgrounds, we provide evidence that the rDNA locus of chromosome 5 is a major source of variation and that the altered 5S rDNA locus position in L*er* impacts chromatin organization in nuclear space. Finally, we suggest that the more permissive chromatin environment at the 5S rDNA locus on chromosome 4 in L*er* reflects differential usage of 5S rDNA loci in different ecotypes.

Conclusions:

The analysis provides evidence for variability in 5S rRNA gene copies and positions in Arabidopsis and suggests a role for chromatin organization in transcriptional regulation and suppression of 5S rDNA translocations.

Keywords:

Tandem repeat, chromatin organization, copy number variation, genome dynamics, nuclear organization

Background

The cell's protein synthesis depends on the ribosome, a highly conserved ribonucleoproteic complex consisting of ~80 ribosomal proteins and 4 different ribosomal RNA (rRNA) molecules. The 5.8S, 18S and 25S rRNAs are transcribed by RNA Polymerase I from conserved clusters, the 45S rDNA loci, while 5S rRNAs are transcribed by RNA polymerase III from 5S rRNA genes. A common characteristic for rRNA genes is their high transcriptional level and their common organization in repeated arrays of nearly identical gene copies due to concerted evolution, a process that promotes homogeneity among the many repeat units (Brown et al., 1972; Dover, 1982; Nei and Rooney, 2005).

In the model plant *Arabidopsis thaliana*, an estimated 700-800 copies of 45S rRNA genes are arranged in head to tail arrays on the tips of the short arms of chromosomes 2 and 4 (Chandrasekhara et al., 2016). The main 5S rDNA loci in Col-0 ecotype are situated in the pericentromeric regions of chromosome 3, 4 and 5 as revealed by Fluorescence *in situ* hybridization (Murata et al., 1997; Fransz et al., 1998) and physical mapping (Tutois et al., 1999). A typical Arabidopsis 5S rRNA gene is 500bp long comprising a 120 bp transcribed sequence and a 380bp spacer region. The transcribed sequence contains an internal promoter region bound by the 5S rDNA-specific transcription factor TFIIIA, and a TATA-like motif, necessary for transcription, resides 28bp upstream of the transcribed region (Cloix et al., 2003; Layat et al., 2012b). Transcription by Polymerase III terminates in a T-rich termination signal, which carries particular sequence motifs, the T-stretch signatures, suggested to be specific for the 5S rDNA copies of the different chromosomes (Cloix et al., 2002). Not all gene copies encode a 5S rRNA transcript identical to the consensus 5S rRNA sequence, instead some accumulated polymorphisms in both transcribed sequence and spacer region (Cloix et al., 2002). Based on these polymorphisms major 5S rRNA genes encoding the consensus 5S transcript are distinguished from minor 5S rRNA genes that carry one to several polymorphisms in the transcribed sequence (Cloix et al., 2002). While in leaf tissue only major genes are expressed, 5S rRNA genes with up to two single nucleotide polymorphisms in the transcribed sequence are expressed in specific tissues, such as the seed, which accumulate high levels of 5S rRNA (Mathieu et al., 2003; Layat et al., 2012a). Built on these observations, the 5S rRNA genes with several polymorphisms are considered to be selectively repressed, but some minor genes retain the potential to be transcribed to suit the demand in protein synthesis in certain cell types or developmental stages. Epigenetic mechanisms have been shown to be involved in selective silencing of certain 45S rRNA

genes (Earley et al., 2006; Lawrence et al., 2004; Pontvianne et al., 2012) and the polymorphic 5S rRNA gene copies (Vaillant et al., 2008; Douet et al., 2009; Mathieu et al., 2003). Indeed, loss of the chromatin remodeling factor DECREASE IN DNA METHYLATION 1 (DDM1) alleviates the repression of minor 5S rRNA gene copies (Vaillant et al., 2006; Mathieu et al., 2003) similarly to mutants in DNA methylation maintenance (*met1*), histone de-acetylation (*hda6*) or the RNA-directed DNA methylation (RdDM) pathway (*ago4*) (Vaillant et al., 2007). The importance of epigenetic control in 5S rRNA gene transcription is also revealed by the expression of an atypical 5S rRNA transcript of 210 bases, which extends beyond the termination sequence into the intergenic spacer region, in different chromatin mutant contexts (Blevins et al., 2009; Vaillant et al., 2006, 2007). Furthermore, specific modes of epigenetic regulations may operate at the different 5S rRNA loci, exemplified by the role of RNA polymerase V in controlling transcription and chromatin organization of the 5S rRNA gene copies situated on chromosome 4 (Douet et al., 2009). However, so far detailed information on potential differences in epigenetic marks at the different 5S rRNA gene loci and their role in the control of 5S rRNA gene transcription are missing.

Indeed, in depth epigenetic analyses have been hampered by the lack of precise reference sequences of the different 5S rRNA gene loci. While ~119 Mb of the Arabidopsis Columbia (Col-0) reference genome was assembled sixteen years ago (Arabidopsis Genome Initiative, 2000), its total genome size has been assessed by flow cytometry as 146 Mb (Hosouchi et al., 2002), 157 Mb (Bennett et al., 2003; Davison et al., 2007), 166 Mb (Long et al., 2013) or even 202 Mb (Schmuths et al., 2004) suggesting that large arrays of heterochromatic repeats are lacking or remain incompletely assembled. These large gaps in the reference genome assembly include the pericentromeric regions that comprise the 5S rRNA gene arrays (Cloix et al., 2000; Fransz et al., 1998; Murata et al., 1997). Furthermore, the arrangement of rRNA genes in tandem-repetitive arrays impeded a detailed analysis. The structure of the 5S rRNA gene loci and the arrangement of 5S rRNA genes including their polymorphisms is therefore still sparse, and the available information is mostly derived from a small number of sequences (Tutois et al., 2002; Campell et al., 1992; Cloix et al., 2000, 2002).

Tandemly repeated arrays such as the 5S rRNA genes have been found to be evolutionary dynamic and to undergo important variation in copy number due to repeat expansion and contraction (Eickbush and Eickbush, 2007). In plants, the copy number of 5S rRNA genes varies extensively, with up to 75 000 copies in certain plant species, however even within the same species copy number variations have been reported (Sastri et al., 1992; Lagudah et

al., 1989). In the Arabidopsis Col-0 genome 5S rRNA gene copies were estimated to about 1000 (Campell et al., 1992) per haploid genome.

Using next-generation sequencing datasets, we show here that the Col-0 genome comprises over 2000 5S rRNA gene copies distributed in three major loci in the pericentromeric regions of chromosome 3, 4 and 5. Each locus contains predominantly 5S rRNA genes carrying a chromosome-specific termination sequence, the T-stretch signature. The 5S rRNA gene sequences on chromosome 3 are highly polymorphic and enriched in repressive histone marks. In contrast, the 5S rRNA genes of chromosome 4 and 5 comprise up to 40% of gene copies identical to the consensus sequence and are moderately enriched in transcriptionally permissive marks. The analysis of a large number of ecotypes further revealed extensive copy number variations between Arabidopsis accessions from different locations. Using linkage mapping in the MAGIC population, chromosome-specific FISH and qPCR as well as *in silico* analysis we showed that variations in 5S rDNA locus organization in the L*er* ecotype as well as in several mutants in the RdDM pathway involves the 5S rRNA gene copies with the chromosome 5-specific T-stretch signature. Finally, we show that a 5S rRNA gene locus at a novel position impacts chromatin organization in interphase nuclei and reveal different enrichment in epigenetic marks between Col-0 and L*er* a the chromosome 4 locus, which suggests differential usage of 5S rRNA genes in distinct ecotypes.

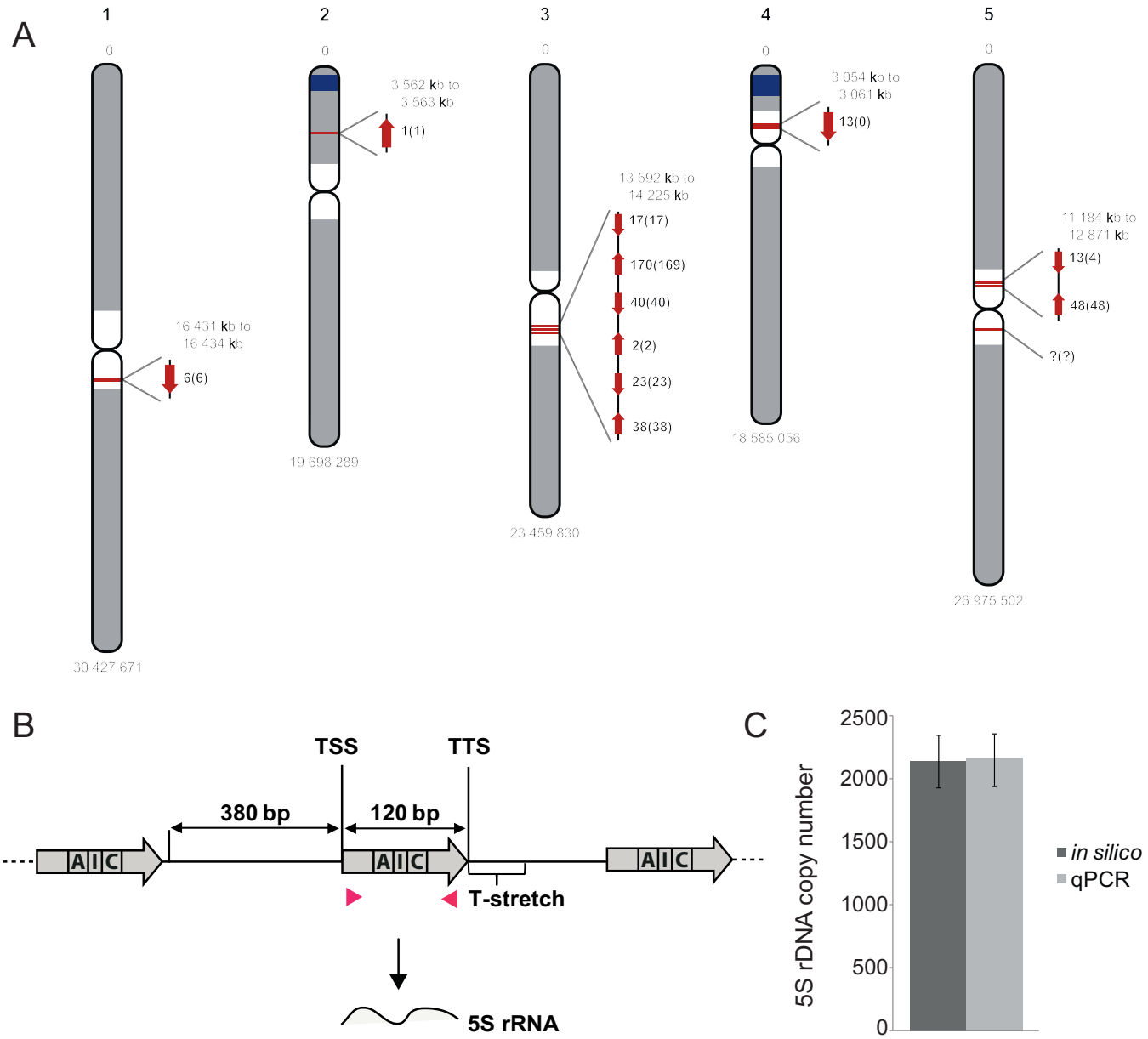


Figure 1: Copy number of 5S rRNA genes in the Columbia (Col-0) ecotype

(A) Schematic representation of the 5 Arabidopsis chromosomes showing the number and orientation of sequences with homology to the 5S rRNA consensus sequence as present in the Tair10 Col-0 genome assembly. The total number of copies is given, and in brackets the number of copies bearing polymorphisms compared to the consensus sequence of the 120bp long transcribed sequence.

(B) Scheme of 2 genes of the 5S rRNA gene array, comprising the 120bp transcribed sequence (arrow) containing the internal promoter sequence (box A, I, E), the 380bp long intergenic region and the Thymidine-rich termination sequence (T-stretch). Positions of primers (arrowheads) used to determine 5S rRNA gene copy number by qPCR are indicated.

(C) Copy number determined by in silico analysis (dark grey) in the Mi-seq dataset and two available Illumina data sets, and by qPCR (light grey). 5S rRNA gene copy number was determined by normalizing to the reads mapping to the transcribed sequence of the 5S rRNA gene with the normalized reads for 18 single copy genes. For the latter mean values of 5S rDNA copy numbers for three biological replicates normalized to two single copy genes (*HXK1*, At4g29130 and *UEV1C*, At2g36060) are shown.

Results

5S rRNA gene copy number estimation in the Col-0 genome

To assemble the available sequence information, we first performed a BLAST analysis to determine the number and position of 5S rRNA genes in the Arabidopsis genome using the TAIR10 pseudomolecule (Schneeberger et al., 2011). We found 371 sequences with homology to the 5S rRNA consensus sequence (Campell et al., 1992) out of which 93.8% are polymorphic in respect to the consensus of the 120 bp transcribed sequence. The majority of the copies are present in the pericentromeric region of the long arm of chromosome 3 (290) and on the upper arm of chromosome 5 (52, Figure 1A). This number of 5S rRNA genes in the Col-0 reference genome is substantially lower than the 1,000 copies estimated previously (Campell et al., 1992). To more precisely determine the copy number of 5S rRNA genes in the Col-0 genome, we used high coverage sequencing data. We generated an Illumina Mi-seq dataset with 300 bp paired-end reads and fragments of roughly 500 bp in size (74.5% reads >QC 30). We further used two available Hi-Seq datasets for Col-0 from the 1001 genome project and the mutation accumulation (MA) lines (Supplementary Table 1) (The 1001 Genomes Consortium, 2016; Shaw et al., 2000; Hagmann et al., 2015). Read mapping to the consensus of the 120 bp transcribed sequence of the 5S rRNA gene (Figure 1B, Figure S1) and subsequent normalization to a selection of 18 single copy genes, revealed the presence of about 2000 5S rRNA gene copies per haploid genome (Figure 1C). A complementary qPCR-based approach using primer sets mapping to the 120 bp-conserved region (Figure 1B) as well revealed about 2,000 copies (Figure 1C), even though this technique might underestimate copy number due to the presence of highly polymorphic 5S rRNA gene copies.

Specific DNA signatures in the T-stretch downstream of the transcribed sequence distinguish the 5S rDNA loci in the Col-0 genome

To obtain a higher-resolution view of the different T-stretch signatures present in the three 5S rRNA gene loci (Figure 2A) in the Col-0 genome we extracted all sequences 3' of the 120 bp transcribed sequence using the same Illumina data sets and classed them according to the previously identified T-stretch signatures (Cloix et al., 2002) (Figure 2B). T-stretch sequences could be identified as belonging to chromosome 3 by the specific stretch characterized by CGG(Nx)CTC, to chromosome 4 with the uninterrupted T-stretch and to the upper arm of chromosome 5 with the motif ATG(Nx)AACC (Figure 2B). This analysis revealed no

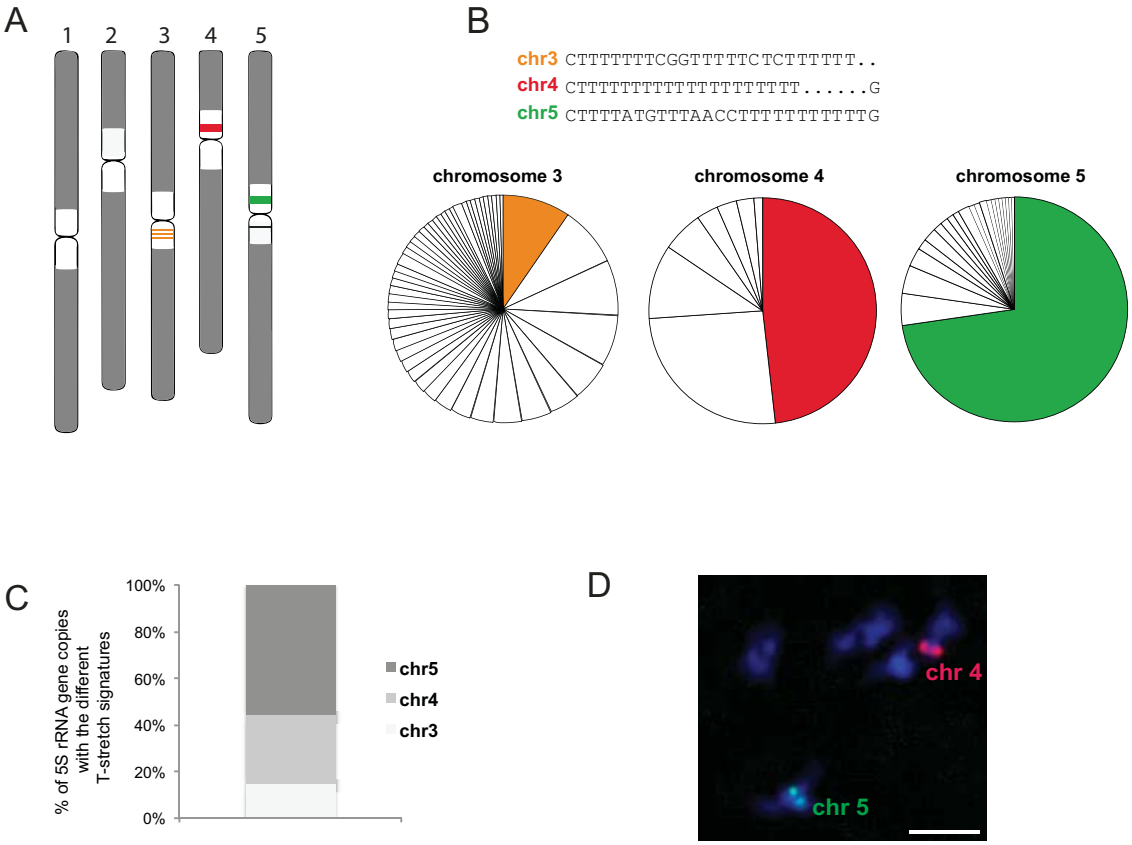


Figure 2 A-D: Characterization of 5S rRNA genes and their genomic distribution in Col-0
(A) Schematic representation of the predominant 5S rDNA loci in the pericentromeric regions (white) of chromosome 3 (orange), 4 (red) and 5 (green) in the Col-0 ecotype .
(B) Principal 5S rDNA signatures (T-stretch) of each 5S rDNA *locus* in the Col-0 genome derived from the Mi-seq data set. Relative frequencies of the three main signatures are indicated as colored fractions in the above pie charts. A more exhaustive analysis of the different polymorphisms found in the T-stretch signatures is given in Supplementary figure 2.
(C) Ratio of 5S rRNA gene copies with the respective T-stretch signature in the Col-0 genome.
(D) DNA FISH on metaphase chromosomes with LNA/DNA mixmer probes designed to specifically recognize the predominant T-stretch signature of chromosome 4 (red) and chromosome 5 (green), DNA is counterstained with DAPI (blue). The scale bar presents 10 µM.

previously unknown T-stretch signatures and also did not identify any T-stretch signature specific to the highly polymorphic 5S rDNA copies on the lower arm of chromosome 5, in agreement with its highly mutated nature and the presence of truncated copies (Cloix et al., 2000). The same sequences as described in (Cloix et al., 2002) were identified as principal DNA-signatures for the chromosome 4 and 5 locus (Figure 2B, Figure S2A), but also recurrent single nucleotide polymorphisms to the predominant signatures were found. For chromosome 4 and 5 these polymorphisms mainly consist of variations in the number of Thymines, while the signature from the chromosome 3 displays most polymorphisms. Among these reads we established the percentage of reads corresponding to 5S rDNA with chromosome 3, 4 or 5-specific T-stretch signatures (Figure 2C). Contrary to the information from the pseudomolecule published (Arabidopsis Genome Initiative, 2000), most 5S rRNA gene copies carry the chromosome 5 signature, followed by those with the signature of chromosome 4 and then of chromosome 3. This is in agreement with previous FISH experiments establishing the locus of chromosome 5 as the largest locus (Murata et al., 1997; Fransz et al., 1998). To confirm that the repeats with the same T-stretch signature are indeed present predominantly at the same genomic position, we analyzed their localization by FISH. We designed LNA-DNA mixmer probes specific to the T-stretch signatures of chromosome 4 and 5. Under optimal hybridization conditions, these probes are highly specific and able to distinguish sequences with only few single nucleotide polymorphisms (You et al., 2006). LNA probes specifically label the 5S rDNA locus of chromosome 4 or 5, respectively (Figure 2D), confirming that the 5S rRNA gene copies with the same T-stretch signature cluster within the same locus.

Identification of chromosome-specific polymorphisms in the transcribed sequence

To determine the polymorphisms within the 120 bp transcribed sequence in a locus-specific manner, we partitioned the reads in three groups according to their specific T-stretch signatures. All reads that, in addition to the T-stretch signature, also covered the complete 120 bp of the consensus sequence of the 5S transcript were mapped and the percentage of copies with 0, 1, 2 or more than 2 SNP and indels determined (Figure 2E). On chromosome 3 nearly all 5S rRNA gene copies carry more than two polymorphisms. The loci from chromosome 4 and 5 comprise, respectively, 43.4% and 33.8% of 5S rRNA genes, which are identical to the consensus sequence, and more polymorphic copies are detected on chromosome 5 compared to chromosome 4. Chromosome 4 and 5 are therefore at the origin of the pool of 5S rRNA genes transcribed in the aerial parts of the plants, in which

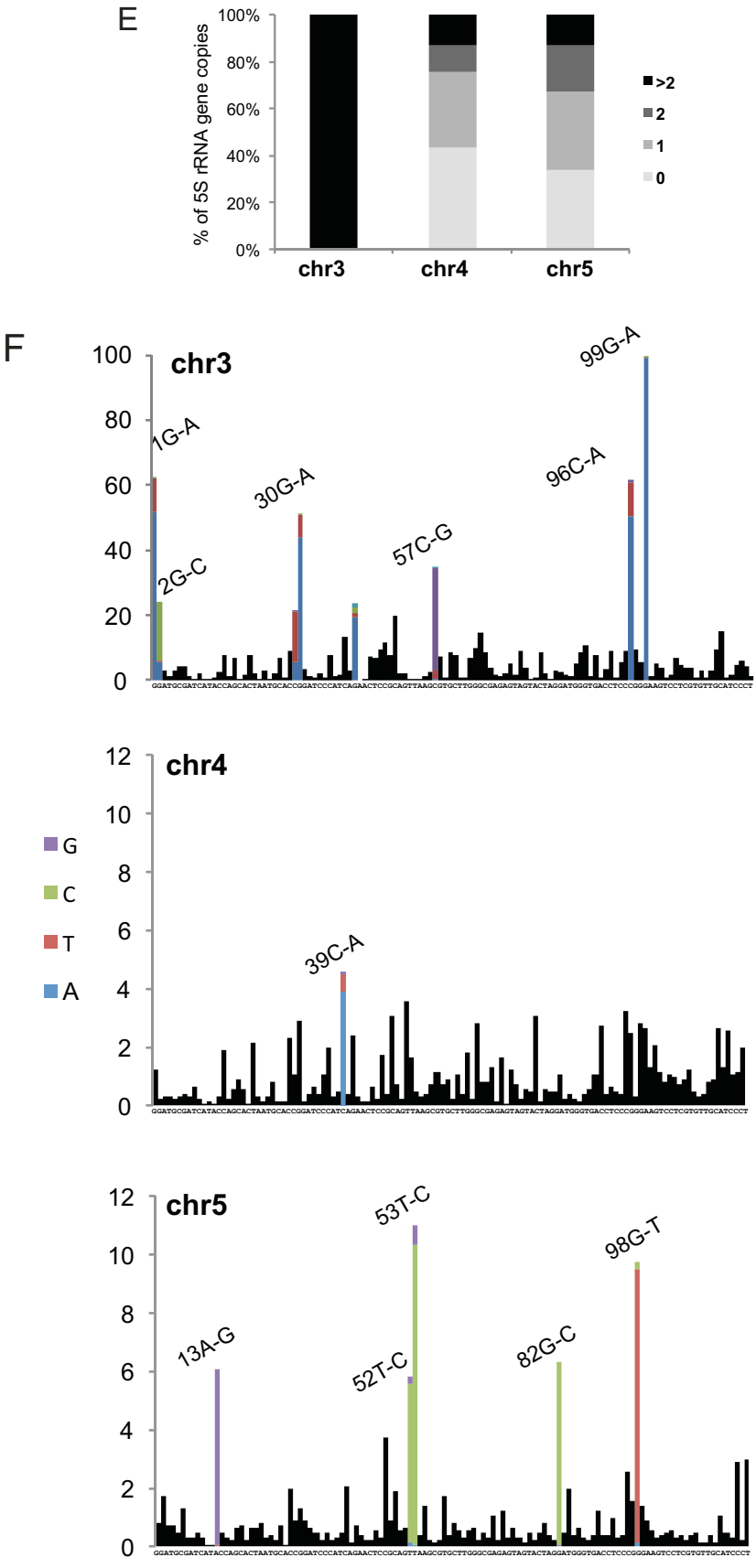


Figure 2 E-F : Characterization of 5S rRNA genes and their genomic distribution in Col-0
(E) Percentage of 5S rRNA gene copies on chromosome 3, 4 and 5 with 0, 1, 2 or more than 2 polymorphisms in the transcribed sequence relative the 5S rRNA consensus sequence.
(F) Percentage of Mi-seq reads comprising the complete transcribed sequence of the 5S rRNA gene that show a specific single nucleotide polymorphisms along the 120 bp-long transcribed sequence per chromosome. The most frequent SNPs as well as the exchanged nucleotide at the given position are indicated in color.

predominantly transcripts without polymorphisms have been detected (Mathieu et al., 2003). A closer investigation of the different single nucleotide polymorphisms revealed some that are specifically or preferentially found on chromosome 5, such as 53 T-C (Figure 2F). SNPs 30 G-A and 99 G-A are present in over 50% and 99% of the 5S rRNA genes on chromosome 3, respectively. To determine whether these SNPs are typical signatures of the chromosome 3 locus in other natural accessions as well, we made use of an experimental population that combines 19 parents (Gan et al., 2011) from a wide geographical distribution, the Multiparent Advanced Generation Inter-Cross (MAGIC) population (Kover et al., 2009). Indeed, the polymorphisms 30 G-A, 41 G-A, 96 C-A and 99 G-A map to the pericentromeric region of chromosome 3 and derive exclusively from the Bur-0, Can-0, Col-0 and Edi-0 founder ecotypes (Figure S2B). Incidentally, those are the only 4 of the analyzed 19 founder lines for which we found chromosome 3 specific T-stretch signatures originally identified in Col-0, which suggests that these specific signatures tend to be part of the same haplotype in chromosome 3 in other accessions as well.

We then extended our analysis of the Col-0 Mi-Seq reads to the polymorphisms in the complete 5S rRNA gene sequence and established a consensus 5S rRNA gene reference sequence based on all extracted sequences per chromosome (Figure S2C). The reference sequence for chromosome 5 is most similar to the 5S rRNA gene sequence previously determined (Campell et al., 1992) in agreement with the fact that the 5S rRNA genes with the chromosome 5-specific signature are the most abundant in the Col-0 genome.

5S rRNA gene loci are differentially enriched in certain histone marks and histone variants

Epigenetic regulation is assumed to selectively silence 5S rRNA genes, in particular the polymorphic copies on chromosome 4 and 5, which include more than 2 SNPs (Layat et al., 2012b). We therefore exploited available ChIP-seq datasets (Supplementary Table 2) to investigate nucleosome occupancy (H3-ChIP, MNase), enrichment in histone modifications (H3K4me3, H3K56ac, H3K36me3, H3K27me1 and H3K9me2) and the presence of canonical histones (H3.1, H2A) or histone variants (H3.3, H2A.W, H2A.X and H2A.Z) at the 5S rRNA genes of the different loci. Enrichment at all 5S rRNA genes was first determined by the number of reads mapping to the transcribed sequence of the 5S rRNA gene normalized to the reads in three Col-0 ChIP input DNA sequencing datasets (Figure 3A, Figure S3A). Enrichment at 5S rRNA genes was compared to two active genes *HEXOKINASE* (*HXK1*) and *ACTIN2* (*ACT2*) with different expression levels as well as the silent Ta3 retrotransposon

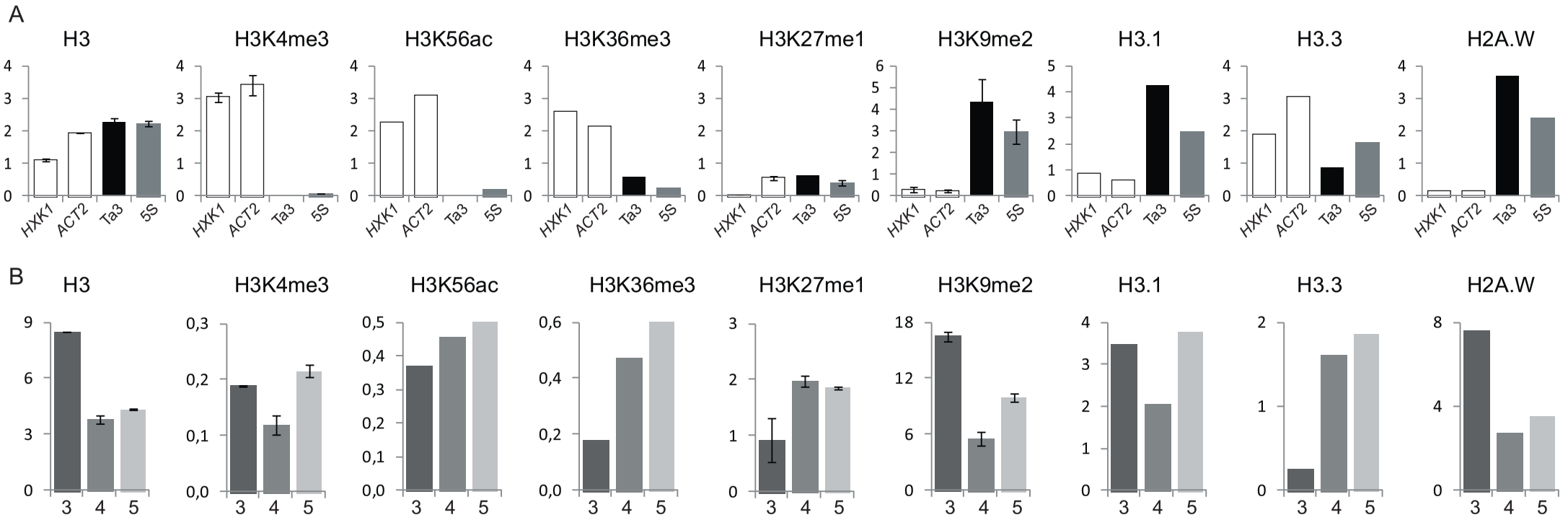


Figure 3 : Differential enrichment in epigenetic marks at 5S rDNA repeats of the different loci

(A-B) Nucleosome occupancy estimated by H3-ChIP and enrichment in histone-post-translational modifications (H3K4me3, H3K56ac, H3K36me3, H3K27me1 and H3K9me2) as well as canonical histones (H3.1) and histone variants (H3.3 and H2A.W) at (A) the transcriptionally active genes *HEXOKINASE 1* (*HXK1*, At4g29130) and *ACTIN2* (*ACT2*, At3g18780), the retrotransposon Ta3 and the 120 bp transcribed sequence of a 5S rRNA gene or (B) at the 3 different 5S rDNA loci. The y-axis corresponds to the number of reads in the ChIP-seq datasets normalized to the number of reads in 3 Columbia ChIP input DNA datasets (see Material and Methods). Error bars corresponding to SEM are shown for those graphs for which more than one dataset was available.

(Figure S3C), situated in the pericentromeric region of chromosome 1. Compared to active genes and similar to Ta3, 5S rRNA genes globally show elevated nucleosome occupancy, are enriched in H3K9me2, the canonical histone H3.1 as well as the histone variant H2A.W, all typical markers of heterochromatin (Stroud et al., 2012; Yelagandula et al., 2014). We then determined the differential enrichment at the three 5S rRNA gene loci based on their specific T-stretch signatures. The locus on chromosome 3 that carries predominantly highly polymorphic copies (Figure 2E-F) is particularly densely packed in nucleosomes, which are depleted of H3.3 and enriched in H2A.W and H3K9me2 (Figure 3B). Interestingly, while the chromosome 3 locus is enriched in H3K9me2, chromosome 4 and 5 loci, which contain the potentially active 5S rRNA genes are enriched in another independently deposited repressive mark namely H3K27me1 (Mathieu et al., 2005). A closer inspection of the distribution of the different histone marks associated with transcription revealed that the 5S rRNA genes both from chromosome 4 and 5 are enriched in H3.3, H3K4me3, H3K56ac and H3K36me3 relative to the gene copies from chromosome 3 (Figure 3) with these four marks being slightly more enriched at the 5S rRNA genes from chromosome 5 compared to chromosome 4. Taken as a whole, 5S rRNA gene loci are largely enriched in repressive chromatin marks, but with an increased occurrence of several active marks at chromosome 4 and 5 with respect to chromosome 3.

5S rRNA gene copy number varies in different ecotypes

The availability of large sets of Arabidopsis accessions that are completely sequenced using Illumina short read sequencing (The 1001 Genomes Consortium, 2016) also allowed us to investigate whether 5S rRNA gene copy number or T-stretch signatures vary between Arabidopsis populations from different geographic populations (The 1001 Genomes Consortium, 2016). We identified ecotypes with as few as 241 and as many as 15 000 5S rRNA gene copies, illustrating that 5S rRNA gene copy number varies extensively between Arabidopsis ecotypes (Figure 4A). To investigate whether the copy number changes between different ecotypes could be related to loss or gain of rRNA gene sequences carrying a specific T-stretch signature that we identified in Col-0, we selected several ecotypes with few (~800), with average (~2000 and ~ 4500) and with the highest copy numbers (up to 15 000), extracted the T-stretch signatures and assigned them to the different loci. The number of 5S rRNA gene copies with signatures both from chromosome 4 and chromosome 5 undergo variations, those from chromosome 5 being the most variable for the ecotypes with up to 5000 copies (Figure 4B). To examine the variation in a shorter time-scale of a single

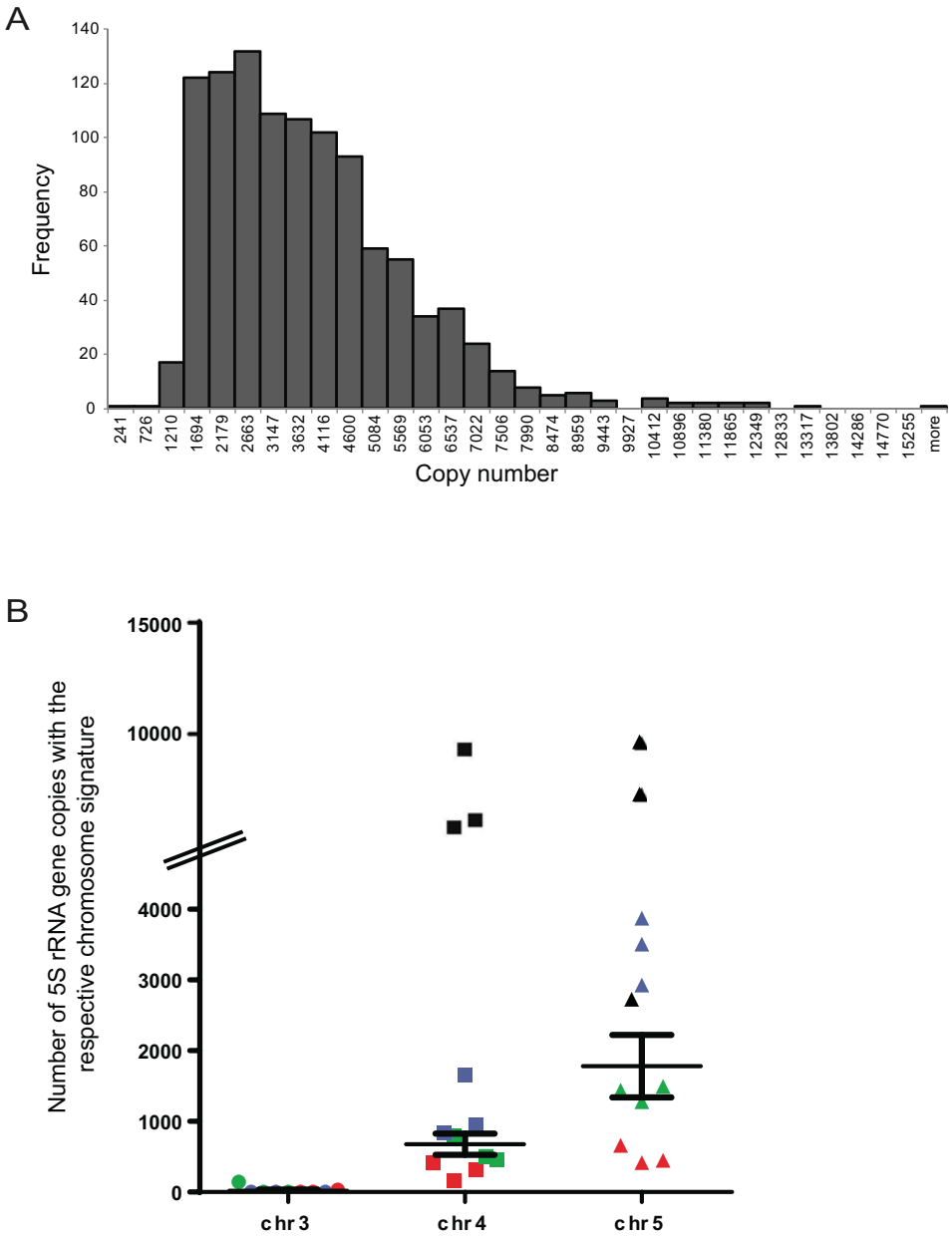


Figure 4: Copy number variation in different *Arabidopsis thaliana* ecotypes
(A) Frequency distribution of 5S rRNA gene copy numbers in 168 different Arabidopsis ecotypes. 5S rRNA gene copy number was determined by standardizing the number of reads mapping to the transcribed sequence of the 5S rRNA gene with the normalized reads for 18 single copy genes.
(B) Number of 5S rRNA gene copies with the respective T-stretches in different ecotypes out of the population in (A) with low (~800; red), middle (~2000 and ~4000; green and blue) and very large (up to 15000; black) 5S rRNA gene copy numbers.

ecotype, we took advantage of the Col-0 mutation accumulation (MA) lines that have been maintained by single-seed descent for 30 generations in the absence of selection (Shaw et al., 2000). We found a rather constant 5S rRNA gene copy number in this population (coefficient of variation = 0.043, Figure S4A), as well as constant relative size of the three different 5S rDNA loci (Figure S4B). This is in sharp contrast to the variation in 45S rRNA gene copy number reported in the same MA population (Rabanal and Nordborg, *unpublished*; coefficient of variation = 0.24).

The additional 5S rRNA gene locus on chromosome 3 in Landsberg *erecta* plants carries the chromosome 5-specific T-stretch signature

To further investigate the variations in 5S rRNA gene loci organization, we analyzed in more detail the Landsberg *erecta* (L*er*) ecotype, the only other ecotype for which a fully independently assembled genome is available (Zapata et al., 2016). The ecotype L*er* further has the interesting particularity to carry a 5S rRNA gene locus on the upper arm of chromosome 3 (Fransz et al., 1998). This additional 5S rRNA gene locus appeared in L*er* plants selected after irradiation from the original Landsberg line (Abraham et al., 2013; Zapata et al., 2016) but the origin of this locus remained unknown. We first determined the T-stretch signatures in available L*er* Illumina sequencing datasets (Abe et al., 2015) and compared these to the ones found in the Col-0 ecotype (Figure S2A). No signature corresponding to chromosome 3 was found in L*er* (Figure S2B) confirming previous observations that noticed neither a 5S rDNA locus on chromosome 3 in Landsberg nor in L*er* large enough to be detected by FISH (Fransz et al., 1998). This suggests that the 5S rRNA gene locus on the upper arm of chromosome 3 in L*er* has a different origin than the chromosome 3 locus in Col-0. We further noticed some variations for chromosome 4 (in the number of thymines) and for the chromosome 5 specific T-stretch signatures, in particular a signature showing the variation T-C and T-G at position 123 and 142 respectively in about 30% of the 5S rRNA genes (Figure S5A). As L*er* is part of the MAGIC population, we used this population to analyze abundant L*er* polymorphisms. We could identify three polymorphisms including 123 T-C (Figure 5A), 142 T–G and 56 G-A (Figure S5B), which all map specifically to the 5S rDNA locus of chromosome 5. Interestingly, all three polymorphisms also map to the upper arm of chromosome 3 with a major QTL at 6.24 Mb according to the Col-0 TAIR10 reference, suggesting that the 5S rDNA locus of chromosome 3 derives from the chromosome 5 locus. Furthermore, *in silico* analysis of the relative abundance of the different T-stretch signatures in Illumina sequencing data sets from La-0

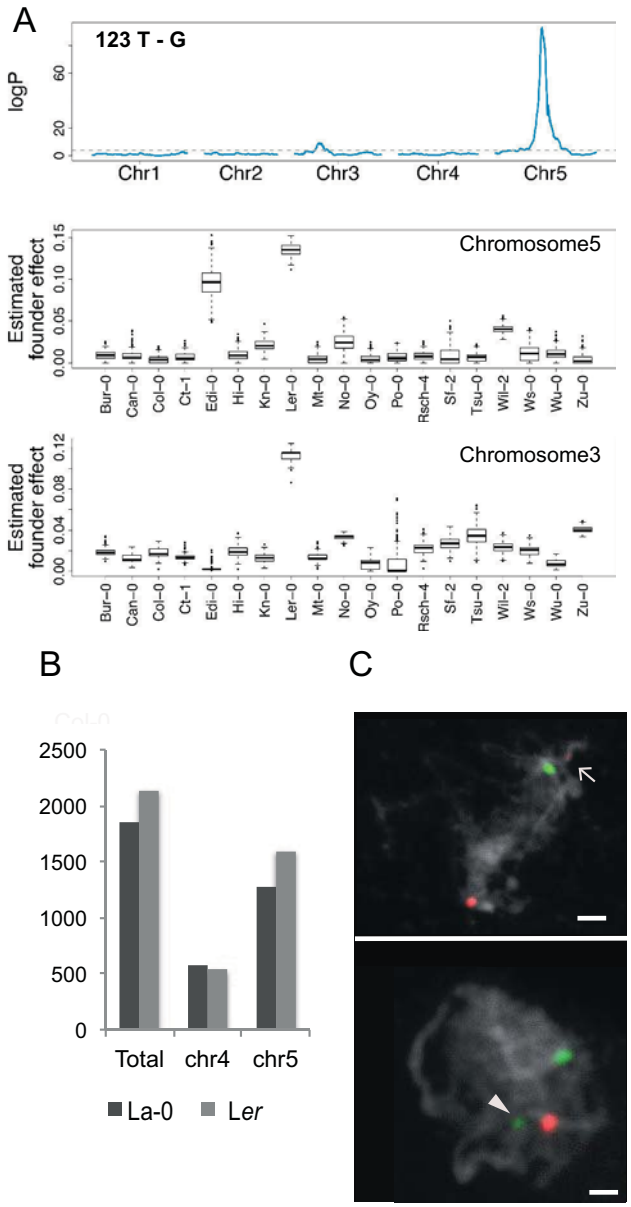


Figure 5 A-C: Variations of 5S rDNA copy number and genomic position

(A) Mapping of the abundance of T to G single nucleotide polymorphism at position 123 estimated by DNA-sequence coverage in 393 individuals of the MAGIC population. Boxplots of the estimated founder accession effect by multiple imputation using R/happy (Mott et al. 2000; Kover et al. 2009) at the positions indicated in the top panel for chromosome 5 and chromosome 3.

(B) Number of total 5S rRNA genes and genes with chromosome 4 or 5 specific T-stretch signatures as determined in La-0 and L*er* Illumina sequencing datasets.

(C_ FISH using LNA/DNA mixmer probes specific for chromosome 4 (red) and chromosome 5 (green) on pachytene spreads of Col-0 and L*er*. DNA is counterstained in grey. Arrow indicates weak cross hybridization of the chromosome 4 probe to the 5S rRNA gene copies of chromosome 3. Arrowhead indicates the locus on the upper arm of chromosome 3 in L*er*.

and L*er*, revealed more 5S rDNA copies in L*er* compared to La-0 and this increase is due to a higher number of 5S rRNA genes with the chromosome 5 specific T-stretch (Figure 5B). Finally, the detection of 5S rDNA loci with locus-specific LNA/DNA mixmer probes clearly identified a second locus with the chromosome 5-specific signature (Figure 5C).

5S rRNA loci organization is altered in certain mutants of the RdDM pathway

5S rRNA genes are highly methylated (Mathieu et al., 2002) in part through the RNA-directed DNA methylation pathway (RdDM) (He et al., 2009; Gao et al., 2010; Vaillant et al., 2007; Blevins et al., 2009). DICER-LIKE (DCL) and AROGONAUTE 4 (AGO4) are key components of the RdDM pathway and loss of AGO4 induces hypomethylation of 5S rRNA genes as well as expression of minor 5S rRNA genes (Vaillant et al., 2007). We therefore investigated whether key regulators of epigenetic control of 5S rRNA genes could impact 5S rDNA stability. Interestingly, metaphase spreads of *dcl2-2* and *dcl3-1* revealed small additional 5S rRNA loci on chromosome 1 and 3 or on chromosome 3 respectively (Figure S5C). Altered 5S rDNA organization was also observed in two *ago4* mutant alleles, namely *ago4-2* in the Col-0 background, as well as *ago4-1* in the L*er* background. In *ago4-2*, an additional 5S rDNA locus appeared on the upper arm of chromosome 3, and in *ago4-1*, the 5S rDNA locus on the upper arm of chromosome 3 in L*er* (Fransz et al., 1998) is absent (Figure 5D). Given that 5S rDNA locus variation occurred in two *ago4* mutant alleles and in two different ecotypes, we quantified 5S rDNA copy number by qPCR analysis. 5S rRNA gene copy number was higher in *ago4-2* compared to Col-0 and 5S rDNA copies are lost in *ago4-1* compared to L*er* (Figure 5E). We then wanted to know whether 5S rRNA genes with a specific T-stretch signature were at the origin of these copy number variations and translocations. Using primers pairs positioned on the T-stretch signatures and which amplify specifically the 5S rRNA gene copies from chromosome 3, 4 or 5 (Figure S5D-G), we determined the copy numbers relative to Col-0. As expected, we did not find any 5S rRNA gene copies with the chromosome 3 specific signatures in L*er* and *ago4-1* (Figure 5F) and revealed more 5S rRNA gene copies with the chromosome 5 signatures in L*er* versus Col-0 confirming the mapping and FISH results. While the qPCR with locus specific primers did not permit to identify the origin of the extra locus in *ago4-2* mutants, FISH with LNA/DNA mixmer probes on pachytene spreads showed that this additional locus comprises the chromosome 5 specific T-stretch (Figure 5G). In the *ago4-1* mutant, FISH revealed the loss of the locus on the upper arm of chromosome 3 that carries the chromosome 5 signature, but qPCR results also suggest that, in addition to the loss of 5S rRNA gene copies with the chromosome 5

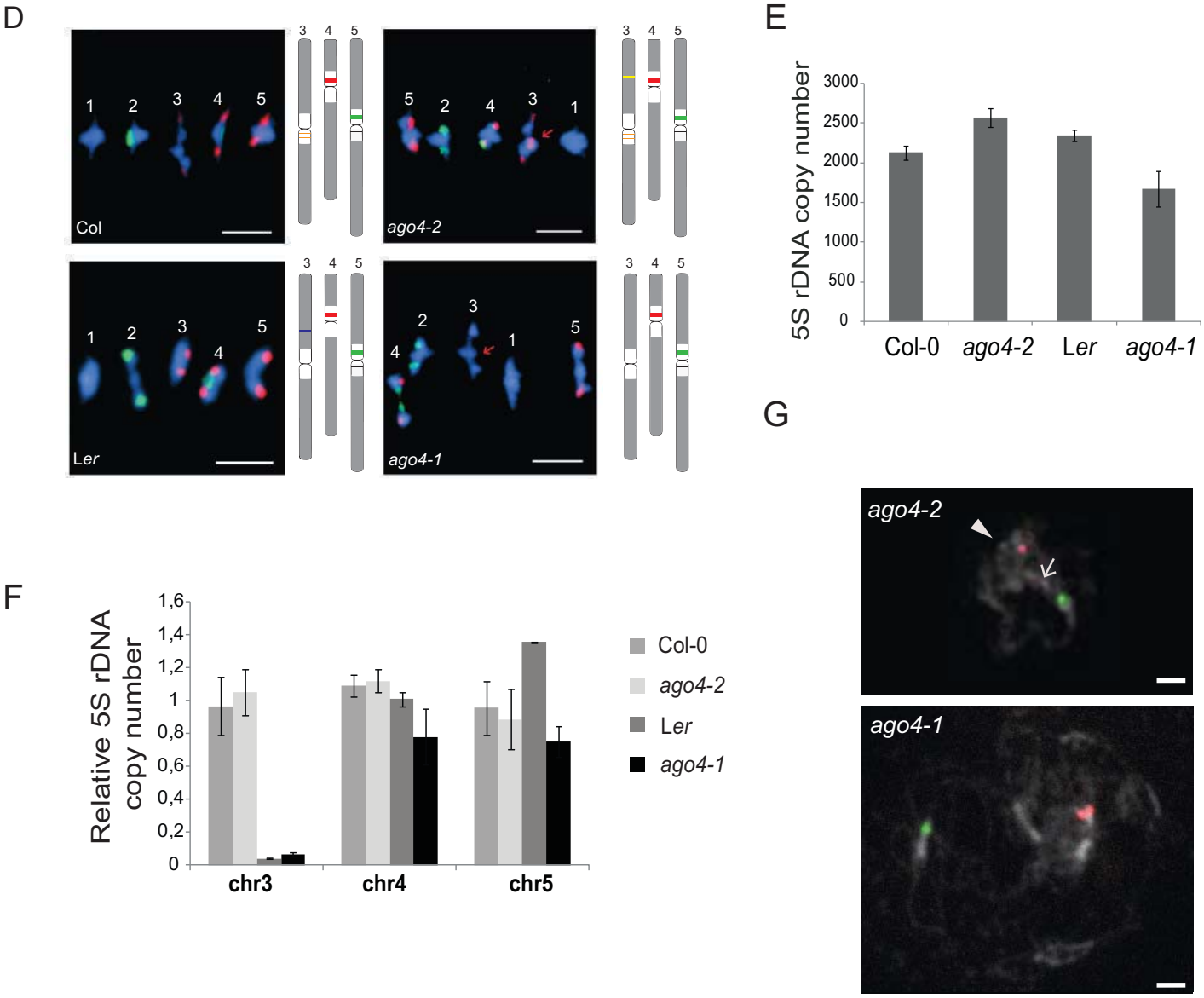


Figure 5 D-G : Variations of 5S rDNA copy number and genomic position

(D) FISH using 5S (red) and 45S (green) probes on metaphase chromosomes of Col-0, L*er* and the ago4 mutant alleles *ago4-2* (in Col-0 background) and *ago4-1* (in L*er* background).

(E) Estimation of 5S rDNA copy number by qPCR in Col-0, *ago4-2*, L*er* and *ago4-1*.

(F) Estimation of relative 5S rDNA copy number by qPCR for each locus in Col-0, *ago4-2*, L*er* and *ago4-1*. Copy numbers in Col-0 are set to 1 for each chromosome.

(G) FISH on *ago4-2* and *ago4-1* pachytene spreads with LNA/DNA mixmer probes. Arrow indicates weak cross hybridization of the chromosome 4 probe with the 5S rRNA gene copies of chromosome 3 in *ago4-2*. Arrowhead indicates the additional locus on the upper arm of chromosome 3 in *ago4-2*.

signature, the chromosome 4 locus might have undergone rearrangements in this mutant (Figure 5F).

Taken together, 5S rRNA loci are dynamically reorganized in several mutants of the RdDM pathway with examples of loss or gain observed in two *ago4* mutant alleles and suggesting that frequent translocations may preferentially initiate from the chromosome 5 locus.

Altered localization of 5S rRNA gene loci impacts nuclear organization and epigenetic marks

Given that 5S rDNA loci show different chromosome positions in Col-0 and L*er* plants, we were interested to know whether this additional 5S rDNA locus on the euchromatic arm of chromosome 3 could affect nuclear organization or epigenetic modifications at 5S rRNA genes. We first carried out FISH on Col-0 and L*er* interphase nuclei using a 5S rDNA probe labeling all 5S rRNA gene loci and the centromeric 180 bp probe to mark chromocenters (Simon et al., 2015). As expected, the majority (~80%) of Col-0 nuclei show six independent signals partly co-localizing with chromocenters (Figure 6A-B). In contrast, in L*er*, ~ 30% and ~20% of interphase nuclei show only 4 or 5 signals, respectively, suggesting frequent co-localization of the 5S rDNA loci on the arm of chromosome 3 with themselves and with the loci present in the pericentromeric regions of chromosome 4 and 5 (Figure 6A-B). Indeed, despite the fact that this locus is physically located far from the centromere, only 20 +/- 9.5% of L*er* nuclei exhibit at least one 5S rDNA signal distant from chromocenters. To distinguish, which loci are involved in the co-localization we scored the number of co-localization events between chromosome 4 and 5 specific hybridization signals using locus-specific LNA/DNA mixmer probes. Co-localization of loci carrying the chromosome 4 and the chromosome 5 signatures was observed in 10% of L*er* nuclei (Figure S6A). Together, this suggests that in some of the interphase nuclei the loci situated on the arms of chromosome 3 induce formation of a chromatin loops to associate with 5S loci in the pericentromeric region and preferentially with the loci situated on chromosome 5.

We then asked whether the position of 5S rRNA copies on the euchromatic part of chromosome 3 might result in a more permissive chromatin environment. To this aim, we carried out ChIP-qPCR in Col-0 and L*er* plants and determined the enrichment in H3K4me3, a mark associated with transcription, and the repressive mark H3K9me2 at all 5S rRNA gene copies and of those situated on chromosome 4 and 5 using locus-specific primers. To our surprise, no loss and even slight enrichment in H3K9me2 was observed at those 5S rRNA

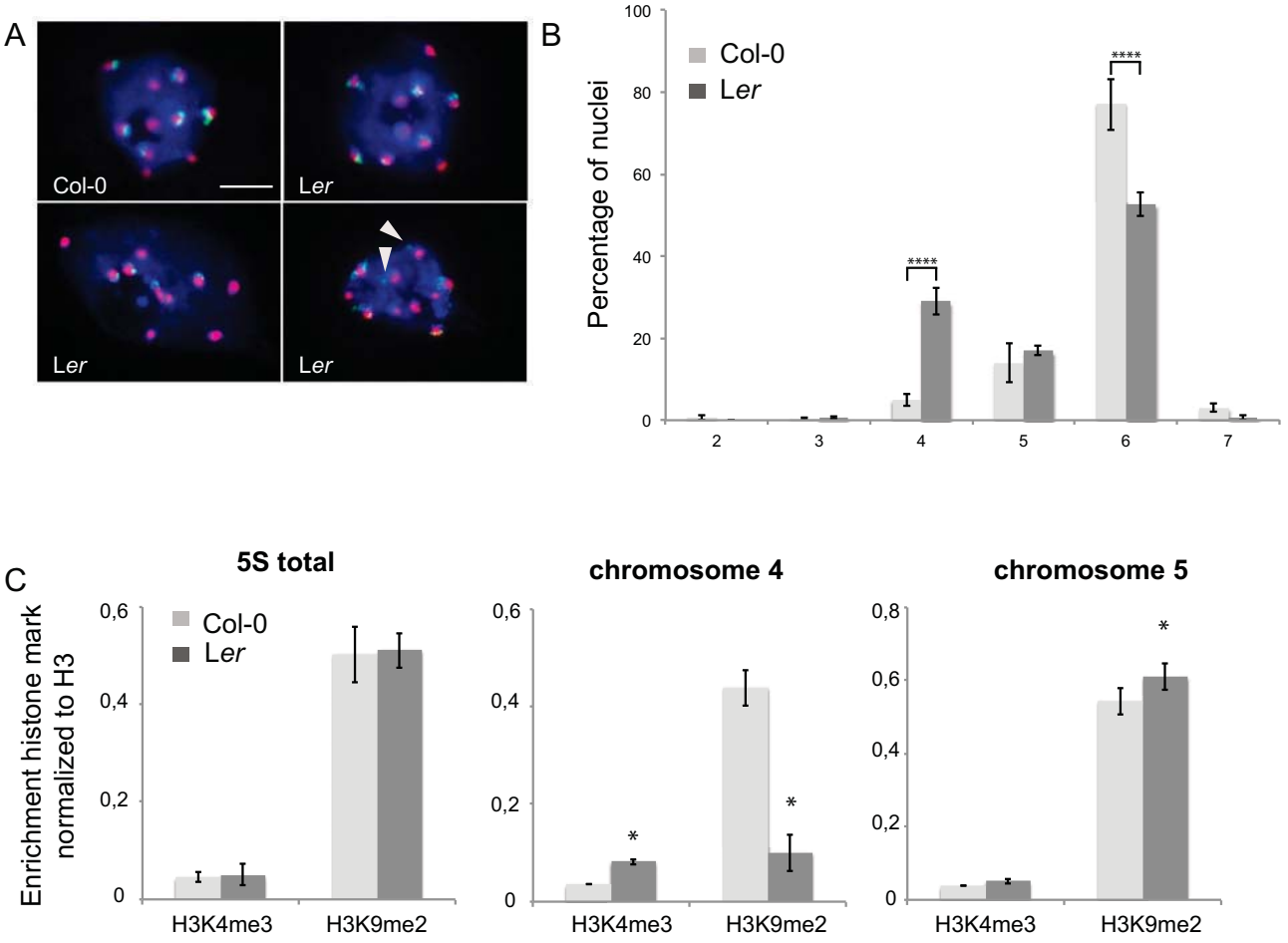


Figure 6: Differences in nuclear organization and epigenetic marks of 5S rDNA loci in Col-0 and L*er*

(A) Representative nuclei of 10 day-old cotyledons stained by FISH with 5S rDNA probes (green) showing 5S *loci* associated or not with a chromocenter revealed by a probe against the centromeric 180bp repeats (red). In Col-0 (left) all 5S signals are associated with chromocenters. Representative nuclei for Ler showing four or six or four 5S rDNA signals associated with chromocenters or six 5S rDNA signals out of which 2 are located distant from chromocenters (indicated by arrowheads).

(B) Percentage of cotyledon interphase nuclei presenting 3, 4, 5, 6 or 7 5S rDNA signals per nucleus as revealed by a FISH probe detecting all 5S rRNA genes (300 nuclei per genotype, n= 3 experiments; Error bars correspond to SEM). ****: < 0.0001 Two-way ANOVA with Sidak's correction.

(C) Differential enrichment in H3K4me3 and H3K9me2 normalized to H3 levels at 5S rDNA loci in Col-0 and Ler plants at all 5S rRNA genes (left), specifically at copies carrying the chromosome 4 (middle) or chromosome 5-specific signature (right). Error bars correspond to SEM from 3 biological replicates. *: <0.05 Wilcoxon test.

genes that carry the chromosome 5 specific signature, while the copies situated on chromosome 4 show a more transcriptionally favorable environment with increased enrichment in H3K4me3 and reduced H3K9me2 levels in L*er* plants. These different epigenetic signatures of the 5S rRNA genes situated on chromosome 4 between Col-0 and L*er* suggest preferential expression of the 5S rRNA gene copies from chromosome 4 in L*er* plants compared to Col-0.

Discussion

Organization of 5S rRNA genes in *Arabidopsis thaliana*

Despite the importance of rRNA in the critical cellular function of protein synthesis still little is understood concerning the organization and regulation of 5S rRNA genes. Their organization in tandemly repetitive arrays as well as their genomic localization in the yet unassembled pericentromeric regions of the Arabidopsis chromosomes renders their study difficult. Here, we provide in depth information on 5S rRNA gene sequences and polymorphisms using available next generation sequencing datasets and an *in house* generated Mi-seq dataset. We confirm by *in silico* analysis and FISH studies with highly specific LNA/DNA mixmer probes that, indeed, 5S rRNA genes on a particular chromosome carry chromosome specific T-stretch signatures.

This information further allowed determining chromosome-specific polymorphisms that are expected to constitute useful information for further epigenomic and transcriptomic analysis of 5S rRNA genes. The 5S rRNA gene cluster are localized on chromosome 3, 4 and 5, with the chromosome 3 locus being the most polymorphic carrying frequent and highly characteristic single nucleotide polymorphisms throughout the transcribed sequence of the 5S rRNA gene, but also in the designated TATA box (Cloix et al., 2003), (Figure S2C). This is in agreement with previous studies suggesting that the 5S rRNA genes situated on chromosome 3 are not transcribed (Cloix et al., 2003) and the fact that this locus is present only in a subset of ecotypes (in only 8 out of the 34 ecotypes for which we determined the T-stretch signature in this study), suggesting that this locus could have been lost during evolution without affecting the cellular 5S rRNA production

Critical nucleotides distinguishing T-stretch signatures from different chromosomes are conserved between the 34 ecotypes that we have investigated in more detail, but not in closely related species such as *Arabis alpina* (data not shown). Despite this conservation in the species Arabidopsis, certain polymorphisms (such as 123 T-G and 142 T-C in the L*er* ecotype) or differences in the number of thymines in the T-stretch, likely due to slippage during replication, are frequent. These observations are in agreement with previous studies on rRNA genes in different species suggesting that the 5S rRNA gene copies of a given locus are highly similar in sequence, but that important differences can be observed between loci (Sastri et al., 1992) suggesting concerted evolution. Unequal cross-over with sister

chromatid exchange dominating over recombination between homologs has been proposed as driving force for evolutionary dynamics of 45S rRNA loci (Eickbush and Eickbush, 2007).

5S rRNA gene copy number variations

45S rDNA copy numbers have been shown to vary substantially between organisms, and even to dynamically change within a few generations in Drosophila (Averbeck and Eickbush, 2005) and *Daphnia obtusa* (McTaggart et al., 2007). Furthermore, 45S rDNA copy numbers have been shown to correlate with genome size in animal and plant species (Prokopowich et al., 2003) and substantial variations in Arabidopsis 45S rDNA copy numbers partly account for genome size variations between ecotypes (Long et al., 2013). Our data revealed that copy numbers of 5S rRNA genes vary between Arabidopsis ecotypes, ranging from about 240 to 15000 copies and large differences can be observed even between geographically close Arabidopsis populations. However, 5S rRNA gene copy numbers might be less dynamic than 45S rRNA genes as comparisons among MA lines show little variation in both absolute copy number and relative abundance of 5S rRNA genes in the different loci (Long et al., 2013; Mozgová et al., 2010; Rabanal and Nordborg, *unpublished*).

Our study also pointed towards methodological difficulties in formally determining copy numbers in particular for short, highly repetitive sequences using next-generation sequencing datasets. Exhaustive testing revealed no correlation between genome coverage and 5S rRNA gene copy number, but showed differences in copy number determination of data sets with different read lengths (Figure S1). For Col-0, we determined a range of 1880 - 2500 5S rRNA gene copies, depending on the dataset, about twice the amount initially determined by Southern Blot analysis and confirmed by our qPCR analysis. However, concerning the copy number variation, in addition to technical differences, we cannot exclude differences in rDNA copy number between Col-0 ecotypes propagated independently for years in different laboratories.

Dynamics of 5S rRNA gene loci

In addition to copy number variation, different genomic locations of 5S rRNA gene loci have been reported in the L*er* ecotype, which was selected after irradiation (Redei, 1962; Fransz et al., 1998) and in different mutants of the RdDM pathway (this study). Independent approaches including mapping using the MAGIC population, *in silico* analysis of available Illumina sequencing datasets, qPCR analysis and FISH studies allowed us to assign the 5S rRNA gene copies on the upper arm of chromosome 3 in L*er* as originating from

chromosome 5. The mapping population further allowed us to narrow down the position of this 5S rDNA locus to ~6.24 Mb, close to a large-scale inversion of 170kb recently found in the L*er* genome, which was likely introduced during the mutagenesis (Zapata et al., 2016). Irradiation of the original Landsberg line might therefore have induced an inversion on chromosome 3 as well as transposition of part of the 5S rDNA locus of chromosome 5. Furthermore, the characterization of *ago4-1* and *ago4-2* mutants, showed loss of the 5S rDNA locus from the upper arm of chromosome 3 and appearance of an additional locus containing 5S rRNA genes with the chromosome 5-specific signature, respectively, illustrating the mobility in particular of the 5S rRNA gene copies from chromosome 5. Together, these observations in the species Arabidopsis support the notion that 5S rRNA genes show frequent reorganization during evolution. Indeed, even the relocation of 5S rRNA genes from the 45S cluster to a separate genomic position or its re-insertion has been observed in different plant species (Garcia and Kovařík, 2013; Wicke et al., 2011). In addition, in Arabidopsis Col-0 plants, the insertion of one complete 45S gene unit into the 5S rRNA gene locus on chromosome 3 was shown (Abou-Ellail et al., 2011). 5S rRNA gene organization may therefore be only under low selective pressure during evolution (Garcia and Kovařík, 2013). What drives these dynamics is still under debate, but the implication of retrotransposons, such as Cassandra elements that comprise 5S rDNA sequences with their LTR region (Kalendar et al., 2008) has been excluded as a possible driving force (Wicke et al., 2011).

Epigenetic regulation of 5S rRNA genes

Despite extensive copy number variations, rRNA genes are assumed to be maintained in excess over the amount required for organism survival (Eickbush and Eickbush, 2007). While these ‘extra’ copies could represent a reserve to be expressed only under certain conditions that require the production of high amounts of rRNA (Layat et al., 2012b, 2012a; Mathieu et al., 2003), a role for silent 45S rRNA gene copies in genome stability has also been discussed (Dvořáčková et al., 2015; Kobayashi, 2011). Given the presence of 5S rRNA genes in excess, a selection of genes undergoes silencing through epigenetic mechanisms. We show that 5S rRNA genes in Col-0 are mostly associated with elevated nucleosomal occupancy and enriched in epigenetic marks associated with heterochromatin (Roudier et al., 2011; Sequeira-Mendes et al., 2014) such as H3.1, H2A.W and H3K9me2. These marks correspond to the repressive chromatin environment and this implies that only a small number of 5S rRNA genes is actively engaged in transcription, similar to the 45S rRNA

genes, for which a range of 10 to 27% are estimated to be transcriptionally active (Pontes et al., 2003; Grummt and Pikaard, 2003). However, most of our knowledge on epigenetic marks in gene regulation is derived from studies on RNA polymerase II transcribed gene loci or transposable elements. Differences to the transcriptional regulation of Pol III transcribed genes may exist, even though recent genome-wide analysis showed that most histone post-translational modifications or histone variants correlate with the same transcriptional response for Pol II and Pol III transcribed genes (Barski et al., 2010; Bhargava, 2013; White, 2011). Active 45S rRNA genes have moreover been suggested to be nucleosome depleted in human cells (Hamperl et al., 2013; Zentner et al., 2011) and therefore 5S rRNA transcription might also occur by creating local nucleosome-free regions without necessarily setting typical active chromatin marks.

Despite the global enrichment in repressive marks, differences can be detected between the three 5S rDNA loci. The copies on chromosome 3 are particularly enriched in H2A.W and H3K9me2 compared to those on chromosome 4 and 5. An interesting observation is the antagonistic enrichment in H3K9me2 or H3K27me1 between chromosome 3 and chromosome 4/5 respectively. H3K9me2 is tightly linked to DNA methylation (Johnson et al., 2002; Tariq et al., 2003), while H3K27me1 is set independently of DNA methylation (Mathieu et al., 2005), which could indicate differential regulation of the 5S rRNA gene copies on the different chromosomes. The pattern in transcriptionally permissive marks is similar for the copies on chromosome 4 and 5, but slightly more enriched at the chromosome 5-specific gene copies. In this respect it is interesting to note that copy numbers of 5S rRNA genes with the chromosome 5-specific signature are highly variable between ecotypes (Figure S4C) and engaged in translocation in *ago4* mutants (Figure 5), in which chromatin organization is affected (Vaillant et al., 2007). It can be speculated that the locus on chromosome 5 is more susceptible to changes in chromatin organization, in comparison to the 5S rRNA genes on chromosome 4 where additional RdDM-independent pathways operate (Douet et al., 2009). It would be interesting to investigate whether translocation of 5S rRNA genes takes place in other mutants affected in chromatin organization and whether the 5S rRNA gene copies on chromosome 5 are particularly susceptible to undergo translocations.

Impact of rRNA gene locus organization on genome function

Frequent translocations and rearrangements of 5S rDNA loci may have substantial impact on genome function, either by disrupting genomic sequences, locally changing chromatin status or altering 3D organization of chromatin within the nucleus. Work from Drosophila illustrated that deletion of 45S rRNA genes affected genome wide gene expression patterns partly as a result of altered heterochromatin composition of the genome (Paredes and Maggert, 2009; Paredes et al., 2011). Here we show that the locus on the upper arm of chromosome 3 in L*er* frequently co-localizes with chromocenters and the other 5S rDNA loci in the pericentromeric regions, in agreement with the preferential clustering of repetitive elements in interphase (Pecinka et al., 2005). The co-localization might thereby introduce differential euchromatic loop formation compared to Col-0 and potentially impact genome expression (Fransz et al., 2002; Feng et al., 2014; Grob et al., 2014). Interestingly, in mouse embryonic stem cells, a transgene carrying a 5S rDNA sequence has been shown to mediate nucleolar association of its associated genomic region, suggesting that 5S rDNA sequences can contribute to nuclear positioning in mammals (Fedoriw et al., 2012).

Given that in L*er* a substantial number of 5S rRNA genes are located in the euchromatic fraction of the genome, we expected less H3K9me2 at the 5S rRNA genes with the chromosome 5-specific T-stretch signature, however H3K9me2 enrichment seems to be unchanged or rather reinforced at these 5S rRNA gene copies. In contrast, we found a more permissive chromatin environment at the 5S rRNA genes of chromosome 4, implying that this locus is preferentially transcribed in the L*er* ecotype. While this has to be experimentally proven, we can speculate that certain 5S rRNA loci are preferentially transcribed in certain ecotypes, similar to the 45S nucleolus organizer regions (Rabanal and Nordborg, *unpublished*). In Arabidopsis, the 5S rRNA locus from chromosome 4 preferentially clusters with the nucleolus organizer regions from chromosome 4 close to the nucleolus (Fransz et al., 2002). Therefore, this locus might be ideally located to facilitate direct transport of 5S rRNA molecules into the nucleolus and foster pre-ribosome particle assembly. An association of 5S rDNA arrays with the nucleolar periphery has also been observed in mouse ES-cells, even though the 5S rDNA locus is located on a different chromosome than the 45S rDNA loci (Fedoriw et al., 2012). L*er* is derived from a knob-less ancestor and therefore does not carry the paracentric inversion present in Col-0, which relocalized part of the pericentromeric domain onto the arm of chromosome 4 forming a heterochromatic knob (Fransz et al., 2000, 2016). While this inversion did not involve the 5S rRNA genes (Fransz et al., 2000), it might have indirectly impacted usage of 5S rRNA genes situated on

chromosome 4 in the knob-containing ecotypes such as Col-0 by reinforcing a transcriptionally silent environment.

Conclusion

5S rRNA is encoded by highly repetitive short gene arrays. Albeit essential for cellular function, the regulatory mechanisms remain poorly understood due to the lack of a reference sequence. Using next generation sequencing data, coupled to molecular and cytogenetic analysis, we show that the Col-0 genome encodes about 2000 5S rRNA gene copies that are grouped into three independent clusters characterized by a specific T-stretch signature and identify locus-specific polymorphisms. This in-depth sequence information provided solid basis for further studies of epigenetic regulation of 5S rRNA genes in a locus-specific manner. The chromosome 3 locus is highly polymorph, present in only few ecotypes and enriched in transcriptionally repressive chromatin marks. Even though embedded in a largely repressive chromatin environment, some active marks are found at the other loci, in particular at 5S rRNA genes situated on chromosome 5 in Col-0 suggesting to mark the actively transcribed copies.

The comparison of different Arabidopsis ecotypes revealed that while T-stretch signatures are conserved between ecotypes, copy numbers are highly variable within the 1001 genome population. The more detailed analysis of the L*er* ecotype as well as different mutants in the RdDM pathway also indicate a role for chromatin organization in suppressing 5S rRNA gene translocations.

In yeast, transcriptional activity at rRNA genes interferes with sister chromatin cohesion at rDNA loci leading to an increased DNA damage sensitivity (Ide et al 2011). Here we show that 5S rRNA locus on chromosome 5 is enriched in epigenetic marks usually associated with transcription activity suggesting that chromosome 5 is actively transcribed. While this remain to be formally demonstrated, transcription activity may be as in yeast correlated with genetic instability and at the origin of translocation originated from chromosome 5 5S rDNA locus. Such genetic instability may be reminiscent of the existence of a more general rDNA amplification system allowing eukaryotic cells to adapt rDNA copy number and hence rRNA production. This mechanism will adapt translation efficiency modulating ribosomes availability to genome size variation.

Acknowledgements

We are grateful to J-P Pichon and H. Duborjal (Biogemma) for valuable advice in library preparation and for performing the high-throughput sequencing of Col-0 DNA. We thank C. Poncet (GENTYANE sequencing facility) for advice and help with DNA shearing using the Covaris technology. We thank P. Arnaud, Y. Bidet, F. Choulet, F. Pontvianne and J. Saez-Vasquez for insightful discussions. This work was supported by doctoral stipend of the Region Auvergne, ANR grant 'Dynam'Het' ANR-11 JSV2 009 01 and 'SINODYN' ANR-12 ISV6 0001. This work was further supported by the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, University Blaise Pascal and the University Auvergne.

Materials and Methods

Plant material

Different Arabidopsis ecotypes as well as mutant *Arabidopsis* lines were obtained from the Nottingham Arabidopsis Stock Center (NASC) and/or were provided by other laboratories. Homozygous mutants *ago4-2* (Columbia background; Agorio & Vera, 2007) and *ago4-1* (N6364, Landsberg *erecta* background) (Zilberman et al., 2003), *dcl2-2 (*SALK_123586) and *dcl3-1* (SALK_005512) were confirmed by PCR-based genotyping. After 2 days of stratification at 4ºC in the dark, plants were grown on soil in a growth chamber under 16-h light/8-h dark cycles at 22ºC. For *in vitro* culture, seeds were sterilized and sown on germination medium containing 0.8% w/v agar, 1% w/v sucrose and Murashige & Skoog salts (M0255; Duchefa Biochemie, Netherlands). After 2 days of stratification at 4ºC at dark, plants were grown under 16-h light/8-h dark cycles at 23ºC.

DNA extraction

The aerial parts of 3 week-old plants were ground to powder and incubated at 65°C in extraction buffer (Tris pH8 0.1M, EDTA pH8 50mM, NaCl 500mM, SDS 1.23%), potassium acetate is then added to a final concentration of 7.45mM before centrifugation at 16000g. The supernatant was added volume to volume into isopropanol, precipitated 10 minutes at -20°C and centrifuged. The pellet was resuspended into Tris-EDTA buffer and ethanol precipitated. The final DNA pellet was resuspended in water.

qPCR conditions

500ng of purified DNA were used per reaction in the Lightcycler 480 (Roche) with the *LightCycler® 480 SYBR Green I Master* (Roche). To amplify the 5S units with general or chromosome specific primers, we used an annealing temperature of 55°C. For the copies situated on chromosome 4 elongation was performed at 60°C; for all the other amplifications at 72°C. To determine the 5S rDNA copy number real primer efficiency of 5S rDNA and two single copy genes (*HXK1*, At4g29130 and *UEV1C*, At2g36060) was determined simultaneously on a serial DNA dilution. Primer sequences can be found in Supplementary Table 3.

ChIP analysis

Chromatin of 3-week old plantlets grown on soil was formaldehyde cross-linked and chromatin immunoprecipitation carried out as previously described (Bowler *et al.*, 2004) with minor modifications: Chromatin was sheared using the Diagenode Bioruptor (10 cycles of

30s ON and 1.5 min OFF) to fragments of about 300 bp. Protein A-coupled Dynabeads (Invitrogen) were used, and the sonicated chromatin was pre-cleared in presence of magnetic beads for 3 h, before immuno-precipitation with anti-H3 antibody (Abcam, ab1791), anti-H3K9me2 (ab1220, Abcam) and anti-H3K4me3 (04-745, Millipore). Enrichment in histone marks was quantified using qPCR (Roche) and normalized to H3 levels.

Fluorescence in situ hybridization

Fluorescence *in situ* hybridization (FISH) was performed on nuclear or pachytene spreads obtained from 10 day-old cotyledons and flower buds fixed in Ethanol-Acetic Acid using directly labeled Locked Nucleic Acid probes (Exiqon): for global 5S "56FAM_CAAGCACGCTTAACTGCGGAGTTCTGAT", specific for the 5S locus of chr4 "TEX615_ACCAAAAAAAAAAAAAAAAAAAGAGGGATG", of chr5 "56FAM_AAAGGTTAAACATAAAAGAGGGATG", and specific for 180bp centromeric repeats "TEX615_GTATGATTGAGTATAAGAACTTAAACCG". Nuclear spreads were obtained as previously described (Probst et al., 2003). Hybridization was performed for 1 hour at 55°C for the locus-specific probes and at 50°C for the global 5S and the 180bp probes. Post-hybridization washes were carried out at 55°C twice in 2x SSC and once in 0.75x SSC. Slides were mounted in Vectashield containing DAPI (Vector laboratories).

The microscope analysis was performed with a Leica DM6000B with an ORCA-Flash4.0 V2 Digital CMOS camera C11440 (HAMAMATSU).

Next-generation Sequencing of Col-0 DNA

DNA from aerial parts of 3 week-old Col-0 plants was extracted using the DNeasy plant Mini Kit (Qiagen) kit and sheared to a size of about 500bp with M220 Focused-ultrasonicator™ (Covaris). Library preparation was performed with the LTP Library Preparation Kit for Illumina® platforms (KAPA) followed by a size selection of around 500bp using AMPure XP (Agencourt) and ligation of adapters (Pentabase). The KAPA library quantification kit for Illumina platforms was then used to quantify the library. Pair-end reads of 2x300bp were obtained using the Illumina Mi-Seq with Mi-Seq Reagent Kit v3.

Bioinformatics analysis

To estimate 5S rRNA gene copy number through next-generation-sequencing (NGS), we divided the average coverage along the 5S gene by the average coverage along 18 single copy genes (At1g13320, At1g58050, At1g59830, At2g28390, At2g32170, At3g01150, At3g53090, At4g26410, At4g27960, At4g33380, At4g34270, At4g38070, At5g08290,

At5g15710, At5g46630, At5g55840 from the set of reference genes for superior transcript normalization (Czechowski et al., 2005), in addition to genes At2g36060, At4g29130 used for qPCR normalization in this study). For each individual analyzed in this study we mapped single-end (SE) reads separately to a single reference consisting of the 120 bp transcribed sequence of the 5S rRNA gene and to the *A. thaliana* TAIR10 reference genome with BWA-MEM (v0.7.8) (Li and Durbin, 2009; Li, 2013). We used Samtools (v0.1.18) to convert file formats (Li et al., 2009), and Sambamba (v0.6.3) to sort and index bam files (Tarasov et al., 2015). We retrieved per-base read depth of the 5S rRNA gene and the 18 single copy genes described above with the function Depthofcoverage from GATK (v3.5) (Van der Auwera et al., 2013).

We noticed that while sequence coverage along the normalization genes augments roughly proportionate to read length, coverage along the 5S rRNA gene raises disproportionally. As a consequence, the longer the read length, the larger the estimate of 5S rRNA gene copies. We report for all Col-0 data sets in this study (Supplementary Table 1) the impact of read length (50 bp and 100 bp) on 5S copy number (Figure S1). Since read lengths of the 1,135 accessions from The 1001 Genomes Consortium are very heterogeneous (ranges from 30 bp to 143 bp; The 1001 Genomes Consortium, 2016) we report 5S rRNA gene copy number only of those accessions for which their reads can be trimmed to 50 bp in length. We trimmed the reads with trimmomatic (Bolger et al., 2014).

To detect polymorphisms along the 5S rRNA gene in the MAGIC population, we first mapped the reads with BWA-MEM as described above to a 148 bp reference 5S rRNA gene plus downstream T-stretch sequence. Then, to obtain the proportion of each polymorphism we employed a pipeline described elsewhere (Rabanal and Nordborg, *unpublished*). Briefly, it counts the proportion of alternative alleles for each reference position with the function variation_strand from the python package pysamstats (v0.24.2; https://github.com/alimanfoo/pysamstats), and a customized version of it that filters out bases with base quality lower than 20 to later evaluate each variable site for its strand bias (SB) score as calculated in Guo et al., 2012 (Guo et al., 2012); positions with a SB score higher than 0.8 were excluded. QTL mapping in MAGIC lines and multiple imputation to determine estimated founder accession effects were performed with R/happy (Mott et al., 2000; Kover et al., 2009).

To determine the sequence of the different T-stretch signatures downstream of the 120 bp transcribed sequence, we generated an *in house* pipeline that maps 50 bp reads using BWA-

ALN (-t 6 -n 30 -k 4 -M 0 -0 -E 0 -e 30 -o 30) against a single reference consisting of the consensus 120 bp transcribed sequence of the 5S rRNA gene prolonged in 3' by 30 Ns and extracts the downstream sequence. The T-stretch sequence for each read was affected manually to its locus of origin based on Cloix et al., 2002. For reads longer than 50 pb, reads were trimmed.

To analyze the polymorphisms in the 5S rRNA genes in a chromosome specific manner, we developed a pipeline that isolates flashed reads from the Mi-seq dataset using grep on an exhaustive list of the identified T-stretches. The isolated reads were then mapped using BWA-MEM to multiple references of 160 bp in length, which comprise the 120 bp transcribed sequence of the 5S rRNA gene as well as 20 nucleotides upstream and downstream; the latter including the specific T-stretches. Only reads covering the whole 120 bp transcribed sequence were retained to determine the polymporphisms in the transcribed sequence. The observed polymorphisms for each reference were extracted and quantified. The same procedure was carried out to determine the polymorphisms along the whole 5S rRNA gene. New consensus sequences for each chromosome cluster were build by modifying the previous reference (Campell et al., 1992) if a polymorphism was present in more than 50% of the reads belonging to a particular locus.

To analyze the enrichment of different epigenetic marks available raw ChIP-seq data (Supplementary Table 2) have been re-analyzed: For the immunoprecipitation (IP) we extracted the number of reads mapped by bwa aln (-k 4 -t 10) to *ACT2*, *HXK1*, Ta3 and the 120 bp transcribed sequence of the 5S rRNA gene divided by the total number of reads. For normalization for the length of the gene of interest or the 5S copy number, this value was divided by the mean read number from three Col-0 input DNA datasets (Stroud et al., 2012; Yelagandula et al., 2014), as most of the ChIP-seq datasets were not provided with an internal input DNA control. To determine the enrichment at each 5S locus, specific T-stretch signatures were extracted as above and normalized to the proportion of 5S reads from a particular chromosome in the input datasets.

References

Abe, M., Kaya, H., Watanabe-Taneda, A., Shibuta, M., Yamaguchi, A., Sakamoto, T., Kurata, T., Ausín, I., Araki, T., and Alonso-Blanco, C. (2015). FE, a phloem-specific Myb-related protein, promotes flowering through transcriptional activation of FLOWERING LOCUS T and FLOWERING LOCUS T INTERACTING PROTEIN 1. Plant J. **83**: 1059–68.

Abou-Ellail, M., Cooke, R., and Sáez-Vásquez, J. (2011). Variations in a team: Major and minor variants of Arabidopsis thaliana rDNA genes. Nucleus **2**: 294–299.

Abraham, M.C., Metheetrairut, C., and Irish, V.F. (2013). Natural Variation Identifies Multiple Loci Controlling Petal Shape and Size in Arabidopsis thaliana. PLoS One **8**.

Agorio, A. and Vera, P. (2007). ARGONAUTE4 is required for resistance to Pseudomonas syringae in Arabidopsis. Plant Cell **19**: 3778–90.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408**: 796–815.

Van der Auwera, G.A. et al. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline.

Averbeck, K.T. and Eickbush, T.H. (2005). Monitoring the mode and tempo of concerted evolution in the Drosophila melanogaster rDNA locus. Genetics **171**: 1837–1846.

Barski, A., Chepelev, I., Liko, D., Cuddapah, S., Fleming, A.B., Birch, J., Cui, K., White, R.J., and Zhao, K. (2010). Pol II and its associated epigenetic marks are present at pol III- transcribed non-coding RNA genes. Nat struct mol biol **17**: 629–634.

Bennett, M.D., Leitch, I.J., Price, H.J., and Johnston, J.S. (2003). Comparisons with Caenorhabditis (~100 Mb) and Drosophila (~175 Mb) using flow cytometry show genome size in Arabidopsis to be ~157 Mb and thus ~25% larger than the Arabidopsis genome initiative estimate of ~125 Mb. Ann. Bot. **91**: 547–557.

Bhargava, P. (2013). Epigenetic regulation of transcription by RNA polymerase III. Biochim. Biophys. Acta **1829**: 1015–25.

Blevins, T., Pontes, O., Pikaard, C.S., and Meins, F. (2009). Heterochromatic siRNAs and DDM1 independently silence aberrant 5S rDNA transcripts in Arabidopsis. PLoS One **4**: e5932.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120.

Brown, D.D., Wensink, P.C., and Jordan, E. (1972). A comparison of the ribosomal DNA's of Xenopus laevis and Xenopus mulleri: the evolution of tandem genes. J. Mol. Biol. **63**: 57–73.

Campell, B.R., Song, Y., Posch, T.E., Cullis, C. a., and Town, C.D. (1992). Sequence and organization of 5S ribosomal RNA-encoding genes of Arabidopsis thaliana. Gene **112**: 225–228.

Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., and Pikaard, C.S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in Arabidopsis. Genes Dev.: 1–14.

Cloix, C., Tutois, S., Mathieu, O., Cuvillier, C., Espagnol, M.C., Picard, G., and Tourmente, S. (2000). Analysis of 5S rDNA arrays in Arabidopsis thaliana: physical mapping and chromosome-specific polymorphisms. Genome Res. **10**: 679–690.

Cloix, C., Tutois, S., Yukawa, Y., Mathieu, O., Cuvillier, C., Espagnol, M.C., Picard, G., and Tourmente, S. (2002). Analysis of the 5S RNA pool in Arabidopsis thaliana: RNAs are heterogeneous and only two of the genomic 5S loci produce mature 5S RNA. Genome Res. **12**: 132–144.

Cloix, C., Yukawa, Y., Tutois, S., Sugiura, M., and Tourmente, S. (2003). In vitro analysis of the sequences required for transcription of the Arabidopsis thaliana 5S rRNA genes. Plant J. **35**: 251–261.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. PLANT Physiol. **139**: 5–17.

Davison, J., Tyagi, A., Comai, L., Pdf, T.P., Biology, B.M.C.P., Article, I., Url, A., Central, P., and Central, B. (2007). Large-scale polymorphism of heterochromatic repeats in the DNA of Arabidopsis thaliana. BMC Plant Biol. **7**: 44.

Douet, J., Tutois, S., and Tourmente, S. (2009). A Pol V-mediated silencing, independent of RNA-directed DNA methylation, applies to 5S rDNA. PLoS Genet. **5**: e1000690.

Dover, G. (1982). Molecular drive: a cohesive mode of species evolution. Nature **299**: 111–117.

Dvořáčková, M., Fojtová, M., and Fajkus, J. (2015). Chromatin dynamics of plant telomeres and ribosomal genes. Plant J.: n/a-n/a.

Earley, K., Lawrence, R.J., Pontes, O., Reuther, R., Enciso, A.J., Silva, M., Neves, N., Gross, M., Viegas, W., and Pikaard, C.S. (2006). Erasure of histone acetylation by Arabidopsis HDA6 mediates large-scale gene silencing in nucleolar dominance. Genes Dev. **20**: 1283–1293.

Eickbush, T.H. and Eickbush, D.G. (2007). Finely orchestrated movements: Evolution of the ribosomal RNA genes. Genetics **175**: 477–485.

Fedoriw, A.M., Starmer, J., Yee, D., and Magnuson, T. (2012). Nucleolar association and transcriptional inhibition through 5S rDNA in mammals. PLoS Genet. **8**: e1002468.

Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E. (2014). Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. Mol. Cell: 1–14.

Fransz, P. et al. (2016). Molecular, genetic and evolutionary analysis of a paracentric inversion in Arabidopsis thaliana. Plant J.

Fransz, P., Armstrong, S., Alonso-Blanco, C., Fischer, T.C., Torres-Ruiz, R.A., and Jones, G. (1998). Cytogenetics for the model system Arabidopsis thaliana. Plant J. **13**: 867–876.

Fransz, P., de Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I. (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. Proc. Natl. Acad. Sci. U. S. A. **99**: 14584–14589.

Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T., and Jones, G.H. (2000). Integrated cytogenetic map of chromosome arm 4S of A. thaliana: structural organization of heterochromatic knob and centromere region. Cell **100**: 367–376.

Gan, X. et al. (2011). Multiple reference genomes and transcriptomes for Arabidopsis thaliana. Nature **477**: 419–423.

Gao, Z. et al. (2010). An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. Nature **465**: 106–9.

Garcia, S. and Kovařík, A. (2013). Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. Heredity (Edinb). **111**: 23–33.

Grob, S., Schmid, M.W., and Grossniklaus, U. (2014). Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. Mol. Cell: 1–16.

Grummt, I. and Pikaard, C.S. (2003). Epigenetic silencing of RNA polymerase I transcription. Nat. Rev. Mol. Cell Biol. **4**: 641–649.

Guo, Y., Li, J., Li, C.-I., Long, J., Samuels, D.C., and Shyr, Y. (2012). The effect of strand bias in Illumina short-read sequencing data. BMC Genomics **13**: 666.

Hagmann, J., Becker, C., Müller, J., Stegle, O., Meyer, R.C., Wang, G., Schneeberger, K., Fitz, J., Altmann, T., Bergelson, J., Borgwardt, K., and Weigel, D. (2015). Century-scale Methylome Stability in a Recently Diverged Arabidopsis thaliana Lineage. PLoS Genet. **11**.

Hamperl, S., Wittner, M., Babl, V., Perez-Fernandez, J., Tschochner, H., and Griesenbeck, J. (2013). Chromatin states at ribosomal DNA loci. Biochim. Biophys. Acta - Gene Regul. Mech. **1829**: 405–417.

He, X.J., Hsu, Y.F., Pontes, O., Zhu, J., Lu, J., Bressan, R.A., Pikaard, C., Wang, C.S., and Zhu, J.K. (2009). NRPD4, a protein related to the RPB4 subunit of RNA polymerase II, is a component of RNA polymerases IV and v and is required for RNA-directed DNA methylation. Genes Dev. **23**: 318–330.

Hosouchi, T., Kumekawa, N., Tsuruoka, H., and Kotani, H. (2002). Physical map-based sizes of the centromeric regions of Arabidopsis thaliana chromosomes 1, 2, and 3. DNA Res. **9**: 117–121.

Johnson, L., Cao, X., and Jacobsen, S. (2002). Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Curr. Biol. **12**: 1360–1367.

Kalendar, R., Tanskanen, J., Chang, W., Antonius, K., Sela, H., Peleg, O., and Schulman, A.H. (2008). Cassandra retrotransposons carry independently transcribed 5S RNA. Proc. Natl. Acad. Sci. U. S. A. **105**: 5833–5838.

Kobayashi, T. (2011). Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. Cell. Mol. Life Sci. **68**: 1395–1403.

Kover, P.X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I.M., Purugganan, M.D., Durrant, C., and Mott, R. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genet. **5**.

Lagudah, E.S., Clarke, B.S., and Appels, R. (1989). Phylogenetic relationships of Triticum tauschii, the D-genome donor to hexaploid wheat. 4. Variation and chromosomal location of 5S DNA. Genome **32**: 1017–25.

Lawrence, R.J., Earley, K., Pontes, O., Silva, M., Chen, Z.J., Neves, N., Viegas, W., and Pikaard, C.S. (2004). A concerted DNA methylation/histone methylation switch regulates rRNA gene dosage control and nucleolar dominance. Mol. Cell **13**: 599–609.

Layat, E., Cotterell, S., Vaillant, I., Yukawa, Y., Tutois, S., and Tourmente, S. (2012a). Transcript levels, alternative splicing and proteolytic cleavage of TFIIIA control 5S rRNA accumulation during Arabidopsis thaliana development. Plant J. **71**: 35–44.

Layat, E., Sáez-Vásquez, J., and Tourmente, S. (2012b). Regulation of Pol I-Transcribed 45S rDNA and Pol III-Transcribed 5S rDNA in Arabidopsis. Plant Cell Physiol. **53**: 267–276.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Prepr. arXiv **0**: 3.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**: 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics **25**: 2078–2079.

Long, Q. et al. (2013). Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat. Genet. **45**: 884–890.

Mathieu, O., Jasencakova, Z., Vaillant, I., Gendrel, A. V, Colot, V., Schubert, I., and Tourmente, S. (2003). Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in Arabidopsis. Plant Cell **15**: 2929–2939.

Mathieu, O., Probst, A. V, and Paszkowski, J. (2005). Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. EMBO J. **24**: 2783–91.

Mathieu, O., Yukawa, Y., Sugiura, M., Picard, G., and Tourmente, S. (2002). 5S rRNA genes expression is not inhibited by DNA methylation in Arabidopsis. Plant J. **29**: 313–323.

McTaggart, S.J., Dudycha, J.L., Omilian, A., and Crease, T.J. (2007). Rates of recombination in the ribosomal DNA of apomictically propagated Daphnia obtusa lines. Genetics **175**: 311–320.

Mott, R., Talbot, C.J., Turri, M.G., Collins, a C., and Flint, J. (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. Proc. Natl. Acad. Sci. U. S. A. **97**: 12649–12654.

Mozgová, I., Mokros, P., and Fajkus, J. (2010). Dysfunction of chromatin assembly factor 1 induces shortening of telomeres and loss of 45S rDNA in Arabidopsis thaliana. Plant Cell **22**: 2768–2780.

Murata, M., Heslop-Harrison, J.S., and Motoyoshi, F. (1997). Physical mapping of the 5S ribosomal RNA genes in Arabidopsis thaliana by multi-color fluorescence in situ hybridization with cosmid clones. Plant J. **12**: 31–37.

Nei, M. and Rooney, A.P. (2005). Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. **39**: 121–52.

Paredes, S., Branco, A.T., Hartl, D.L., Maggert, K.A., and Lemos, B. (2011). Ribosomal dna deletions modulate genome-wide gene expression: “rDNA-sensitive” genes and natural variation. PLoS Genet. **7**: 1–10.

Paredes, S. and Maggert, K.A. (2009). Ribosomal DNA contributes to global chromatin regulation. Proc Natl Acad Sci U S A **106**: 17829–17834.

Pecinka, A., Kato, N., Meister, A., Probst, A. V, Schubert, I., and Lam, E. (2005). Tandem repetitive transgenes and fluorescent chromatin tags alter local interphase chromosome arrangement in Arabidopsis thaliana. J. Cell Sci. **118**: 3751–3758.

Pontes, O., Lawrence, R.J., Neves, N., Silva, M., Lee, J.-H., Chen, Z.J., Viegas, W., and Pikaard, C.S. (2003). Natural variation in nucleolar dominance reveals the relationship between nucleolus organizer chromatin topology and rRNA gene transcription in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **100**: 11418–23.

Pontvianne, F., Blevins, T., Chandrasekhara, C., Feng, W., Stroud, H., Jacobsen, S.E., Michaels, S.D., and Pikaard, C.S. (2012). Histone methyltransferases regulating rRNA gene dose and dosage control in Arabidopsis. Genes Dev. **26**: 945–957.

Probst, A. V., Fransz, P.F., Paszkowski, J., and Mittelsten Scheid, O. (2003). Two means of transcriptional reactivation within heterochromatin. Plant J. **33**: 743–749.

Prokopowich, C.D., Gregory, T.R., and Crease, T.J. (2003). The correlation between rDNA copy number and genome size in eukaryotes. **50**: 48–50.

Redei, G. (1962). Single loci heterosis. Z Vererbungsl **93**: 164–170.

Roudier, F. et al. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J. **30**: 1928–38.

Sastri, D.C., Hilu, K., Appels, R., Lagudah, E.S., Playford, J., and Baum, B.R. (1992). An overview of evolution in plant 5S DNA. Plant Syst. Evol. **183**: 169–181.

Schmuths, H., Meister, A., Horres, R., and Bachmann, K. (2004). Genome size variation among accessions of Arabidopsis thaliana. Ann. Bot. **93**: 317–321.

Schneeberger, K. et al. (2011). Reference-guided assembly of four diverse Arabidopsis thaliana genomes. Proc. Natl. Acad. Sci. U. S. A. **108**: 10249–10254.

Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S.E., Bastolla, U., and Gutierrez, C. (2014). The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. Plant Cell.

Shaw, R.G., Byers, D.L., and Darmo, E. (2000). Spontaneous mutational effects on reproductive traits of Arabidopsis thaliana. Genetics **155**: 369–378.

Simon, L., Voisin, M., Tatout, C., and Probst, A.V. (2015). Structure and function of centromeric and pericentromeric heterochromatin in Arabidopsis thaliana. Front. Plant Sci. **6**: 1–8.

Stroud, H., Otero, S., Desvoyes, B., Ramírez-Parra, E., Jacobsen, S.E., and Gutierrez, C. (2012). Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. Proc. Natl. Acad. Sci. **109**: 5370–5.

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., and Prins, P. (2015). Sambamba: Fast processing of NGS alignment formats. Bioinformatics **31**: 2032–2034.

Tariq, M., Saze, H., Probst, A. V, Lichota, J., Habu, Y., and Paszkowski, J. (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc. Natl. Acad. Sci. U. S. A. **100**: 8823–8827.

The 1001 Genomes Consortium, X. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell **166**: 481–491.

Tutois, S., Cloix, C., Cuvillier, C., Espagnol, M.C., Lafleuriel, J., Picard, G., and Tourmente, S. (1999). Structural analysis and physical mapping of a pericentromeric region of chromosome 5 of Arabidopsis thaliana. Chromosome Res. **7**: 143–156.

Tutois, S., Cloix, C., Mathieu, O., Cuvillier, C., and Tourmente, S. (2002). Analysis of 5S rDNA Loci among Arabidopsis Ecotypes and Subspecies. Genome Lett. **1**: 115–122.

Vaillant, I., Schubert, I., Tourmente, S., and Mathieu, O. (2006). MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. EMBO Rep. **7**: 1273–8.

Vaillant, I., Tutois, S., Cuvillier, C., Schubert, I., and Tourmente, S. (2007). Regulation of Arabidopsis thaliana 5S rRNA Genes. Plant Cell Physiol. **48**: 745–752.

Vaillant, I., Tutois, S., Jasencakova, Z., Douet, J., Schubert, I., and Tourmente, S. (2008). Hypomethylation and hypermethylation of the tandem repetitive 5S rRNA genes in Arabidopsis. Plant J. **54**: 299–309.

White, R.J. (2011). Transcription by RNA Polymerase III: more compolex then we thought. Nat. Rev. Genet. **12**: 459–463.

Wicke, S., Costa, A., Mu??oz, J., and Quandt, D. (2011). Restless 5S: The re-

arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. Mol. Phylogenet. Evol. **61**: 321–332.

Yelagandula, R. et al. (2014). The Histone Variant H2A.W Defines Heterochromatin and Promotes Chromatin Condensation in Arabidopsis. Cell **158**: 98–109.

You, Y., Moreira, B.G., Behlke, M.A., and Owczarzy, R. (2006). Design of LNA probes that improve mismatch discrimination. Nucleic Acids Res. **34**: 1–11.

Zapata, L., Ding, J., Willing, E., Hartwig, B., Bezdan, D., Jiao, W., and Patel, V. (2016). Chromosome-level assembly of Arabidopsis thaliana Ler reveals the extent of translocation and inversion polymorphisms. Proc. Natl. Acad. Sci. **0**.

Zentner, G.E., Saiakhova, A., Manaenkov, P., Adams, M.D., and Scacheri, P.C. (2011). Integrative genomic analysis of human ribosomal DNA. Nucleic Acids Res. **39**: 4949–4960.

Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science **299**: 716–9.

Supplementary Figure legends

Supplementary Figure 1: Influence of read length on 5S rDNA copy number determination in Col-0

5S rRNA gene copy number was estimated in three independent next-generation-sequencing datasets for Col-0 by dividing the average coverage along the transcribed sequence of the 5S gene by the average coverage along 18 single copy genes (see Materials and Methods).

Supplementary Figure 2: T-stretch signatures and polymorphisms in Col-0 5S rRNA gene sequences

(A) Identified T-stretch signatures and their relative percentages per chromosome in the Mi-seq dataset of the Col-0 genome.

(B) Mapping of the abundance of single nucleotide polymorphisms at positions 30, 41, 96 and 99 estimated by DNA-sequence coverage in 393 individuals of the MAGIC population. Boxplots of the estimated founder accession effect by multiple imputation using R/happy (Mott et al., 2000; Kover et al., 2009) at the QTLs indicated in the top panel for chromosome 3.

(C) Alignment of the consensus sequences for rRNA genes of chromosome 3, 4 and 5 with the reference sequence originally determined by Campell et al (Campell et al., 1992) as derived from the Col-0 Mi-seq dataset.

Supplementary Figure 3: Nucleosomal occupancy, H2A, H2A.X and H2A.Z enrichment on 5S rDNA and relative expression level of Ta3, *HXK1* and *ACT2*

(A-B) Nucleosomal occupancy (MNase) and enrichment in H2A, H2A.X and H2A.Z at (A) the transcriptionally active genes *HEXOKINASE1* (*HXK1*, At4g29130) and *ACTIN2* (*ACT2*, At3g18780), the retrotransposon Ta3 and the 120 bp transcribed sequence of a 5S rRNA gene or (B) at the three different 5S rDNA loci.

(C) Ta3, *HXK1* and *ACT2* expression normalized to AtSAND (At2g28390) from two RNAseq datasets (Zhang et al., 2015) of *A. thaliana* shoot tissue.

Supplementary Figure 4: 5S rRNA copy number in Col-0 single seed descendants

(A-B) Total 5S rRNA gene copy number (A) and copies with the chromosome 3, 4 or 5 specific T-stretch signatures (B) as determined in Illumina datasets for 10 individuals of the mutation accumulation (MA) lines.

Supplementary Figure 5: 5S rDNA polymorphisms and variations in L*er* and *dcl* mutants and locus-specific PCR primer sets

(A) Identified T-stretch signatures and their relative percentages per chromosome in the L*er* genome.

(B) Mapping of the abundance of single nucleotide polymorphisms at positions 142 (part of the T-stretch) and 56 estimated by DNA-sequence coverage in 393 individuals of the MAGIC population. Boxplots of the estimated founder accession effect by multiple imputation using R/happy (Mott et al., 2000; Kover et al., 2009) at the QTLs indicated in the top panel for chromosome 5 and chromosome 3.

(C) FISH analysis using 5S (red) and 45S (green) probes on metaphase chromosomes of *dcl2-2* and *dcl3-1* mutants. Arrows indicate additional 5S loci on chromosome 1 and 3 in *dcl2-2* and on chromosome 3 in *dcl3-1* mutant plants.

(D) Schematic representation of two 5S RNA gene units and position of the different primer sets used.

(E) Specificity of the chromosome-specific primer pairs was confirmed by PCR on YAC DNA containing either the loci of chromosome 3, 4 or 5 or genomic DNA as control.

(F, G) Number of 5S rDNA copies estimated by qPCR using different primer pair combinations. Mean values of 5S rDNA copy numbers determined from three biological replicates and normalized to two single copy genes (*HXK1*, At4g29130 and *UEV1C*, At2g3660). Different qPCR conditions have been used for the chromosome 3 and 5 (F) versus chromosome 4 specific primers (G) (see Material and Methods). Note that primer sets that cover a larger region and are not confined only to the highly conserved transcribed sequence underestimate 5S rRNA gene copy numbers.

Supplementary Figure 6: 5S rDNA loci organization in the nucleus and ChIP-qPCR analysis for H3, H3K4me3 and H3K9me2.

(A) Percentage of nuclei showing co-localisation of 5S rDNA loci carrying the chromosome 5 signature with those carrying the chromosome 4 signature in Col-0 and L*er* nuclei (150~200 nuclei by genotype).

B) Enrichment in H3, H3K4me3 and H3K9me2 normalized to input, showing also IP levels in the no-antibody fraction (mock) for Col-0 at all 5S rRNA genes (left) and specifically at two single copy genes (*HXK1*, At4g29130 and *UEV1C*, At2g3660), the retrotransposon Ta3 and the 5S rDNA transcribed sequence confirming enrichments found in ChIP-seq datasets.

Supplementary Figure 7: Polymorphisms in 5S rRNA genes in the L*er ecotype*

Mean percentage of reads with the chromosome 4 or 5 specific T-stretch that show specific single nucleotide polymorphisms (SNPs) along the 120 bp-long transcribed sequence in two L*er* next generation sequencing datasets (Zapata et al., 2016). For the most frequent SNPs and indels, the exchanged nucleotide at the given position is indicated in the respective color. The consensus sequence as determined from Col-0 is indicated below each graph.

Supplementary Table 1:

Next generation sequencing datasets used for Col-0 and L*er* in this study.

Supplementary Table 2:

List and origin of all the ChIP-seq and RNA-seq datasets used in this paper.

Supplementary Table 3:

Primer sequences used in this study.

Supplementary Materials and Methods

PCR on YAC DNA

PCR on YACs named 7E7 (chromosome 3), 9D3 (chromosome 4), 4E4 (chromosome 5) (Creusot et al., 1995) were performed following the recommendation for the GoTaq flexi DNA polymerase (Promega) with 30 cycles of amplification including 30 seconds denaturation at 95°C, 30 seconds hybridization at 57°C and 20 seconds elongation at 72°C.

Analysis of RNA-seq data

Data used for two RNA-seq analyses for leaf tissue (Zhang et al., 2015) were obtained from the NCBI (http://www.ncbi.nlm.nih.gov/geo/browse/). Reads from RNA-Seq libraries were mapped onto the candidate gene sequences as described (Poulet et al., 2016), expressed in Reads Per Kilobase of transcript per Million mapped reads (RPKM) and normalised to *AtSAND* (At2g28390).

Supplementary References

Campell, B.R., Song, Y., Posch, T.E., Cullis, C. a., and Town, C.D. (1992). Sequence and organization of 5S ribosomal RNA-encoding genes of Arabidopsis thaliana. Gene **112**: 225–228.

Creusot, F., Fouilloux, E., Dron, M., Lafleuriel, J., Picard, G., Billault, A., Le Paslier, D., Cohen, D., Chabouté, M.E., and Durr, A. (1995). The CIC library: a large insert YAC library for genome mapping in Arabidopsis thaliana. Plant J. cell Mol. Biol. **8**: 763–770.

Kover, P.X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I.M., Purugganan, M.D., Durrant, C., and Mott, R. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in Arabidopsis thaliana. PLoS Genet. **5**.

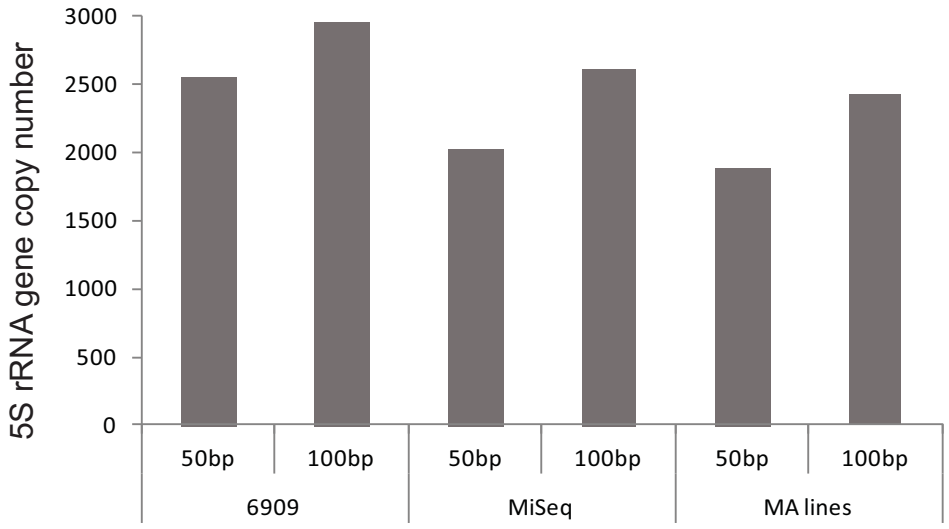
Mott, R., Talbot, C.J., Turri, M.G., Collins, a C., and Flint, J. (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. Proc. Natl. Acad. Sci. U. S. A. **97**: 12649–12654.

Poulet, A., Probst, A. V, Graumann, K., Tatout, C., and Evans, D.E. (2016). Exploring the evolution of the proteins of the plant nuclear envelope. Nucleus: 0.

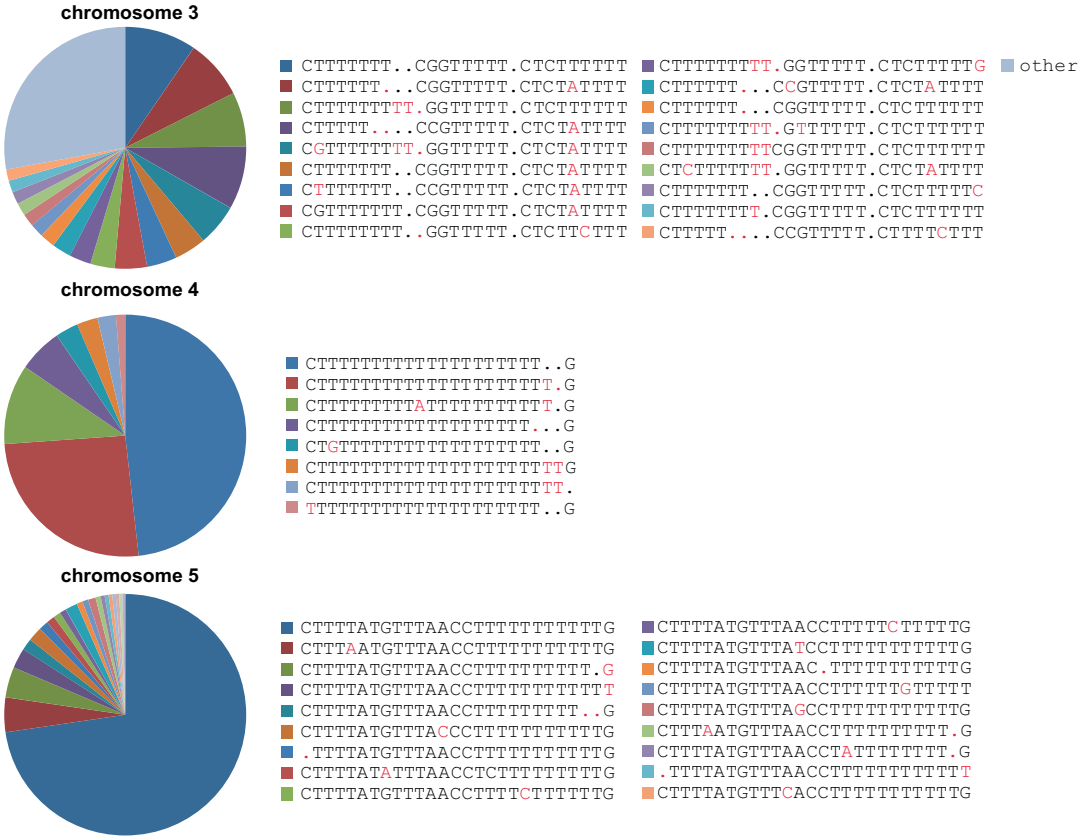
Zapata, L., Ding, J., Willing, E., Hartwig, B., Bezdan, D., Jiao, W., and Patel, V. (2016). Chromosome-level assembly of *Arabidopsis thaliana* L*er* reveals the extent of translocation and inversion polymorphisms. Proc. Natl. Acad. Sci. **0**.

Zhang, X. et al. (2015). Suppression of endogenous gene silencing by bidirectional cytoplasmic RNA decay in Arabidopsis. Science **348**: 120–3.

Supplementary Figure 1



Supplementary Figure 2A



Supplementary Figure 2B

B

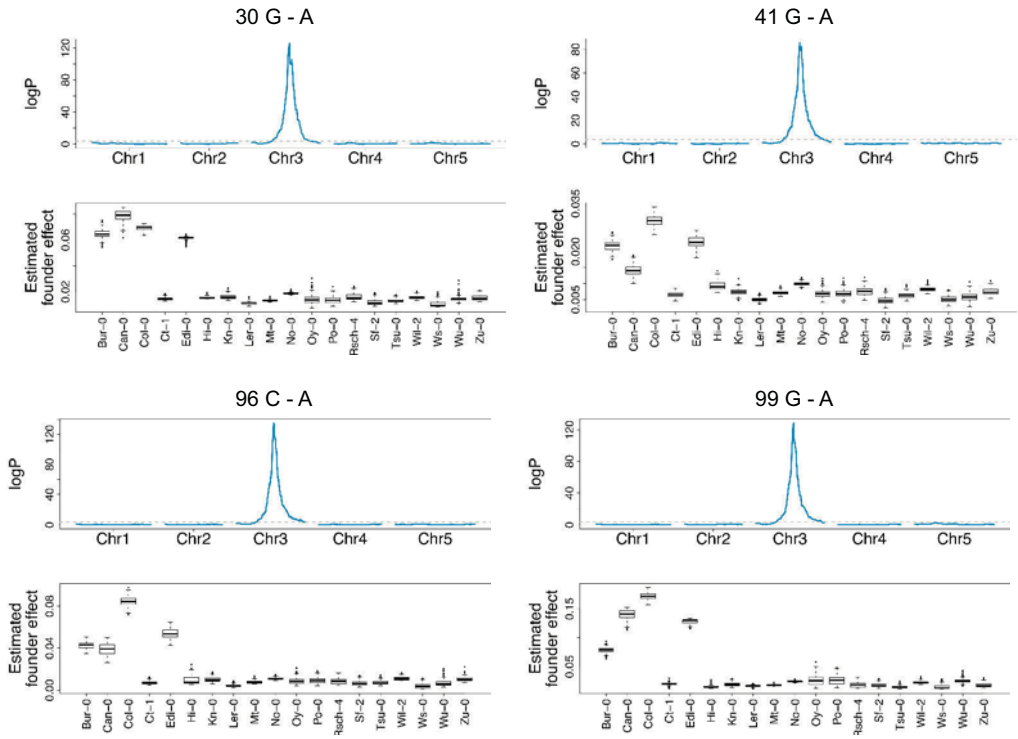
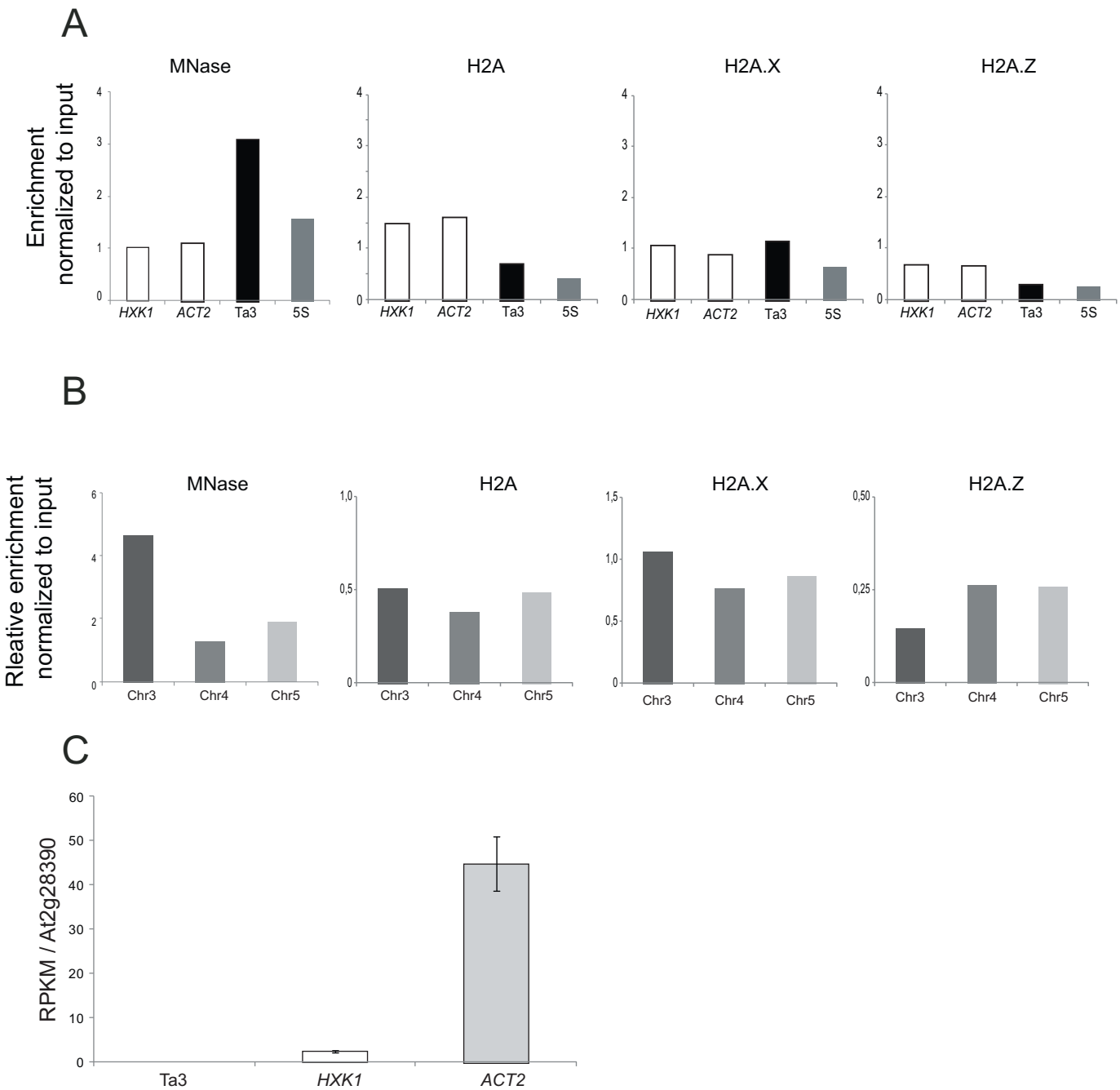


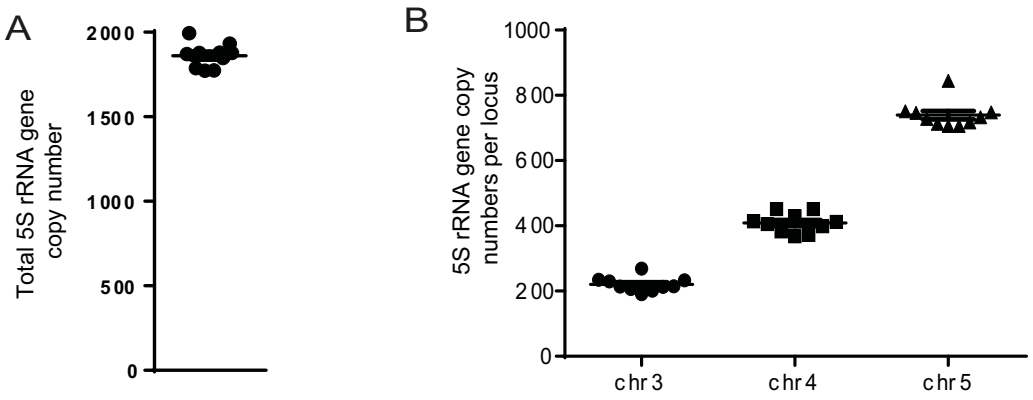
Figure S2C : Alignment of new 5S rRNA gene references.

```
chr3       AGATGCGATCATACCAGCACTAATGCACCGGATCCCATCAGAACTCCGCA
chr4       GGATGCGATCATACCAGCACTAATGCACCGGATCCCATCAGAACTCCGCA
chr5       GGATGCGATCATACCAGCACTAATGCACCGGATCCCATCAGAACTCCGCA
Campell    GGATGCGATCATACCAGCACTAATGCACCGGATCCCATCAGAACTCCGCA

chr3       GTTAAGCGTGCTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCAGGAA
chr4       GTTAAGCGTGCTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCCGGGA
chr5       GTTAAGCGTGCTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCCGGGA
Campell    GTTAAGCGTTCTTGGGCGAGAGTAGTACTAGGATGGGTGACCTCCCGGGA

chr3       AGTCCTCGTGTTGCATCCCTCTTTTTTTCGGTTTTTCTCTTTTTTTTTTG
chr4       AGTCCTCGTGTTGCATCCCTCTTTTTTT-TTTT------TTTTTTTTGG
chr5       AGTCCTCGTGTTGCATCCCTCTTT-TAT-GTTTAACCTTTTTTTTTTGG
Campell    AGTCCTCGTGTTGCATCCCTCTTTATAT-GTTTAACCTTTTTTTTTTGG

chr3       TTAAAAATGTATGACTCTATAACTTTTAGACCGTGAGGCCAAACTTGGCA
chr4       TTAAAACTTTATGACTCTATAACTTCTATACCGTGAGGCCAAACTTGGCA
chr5       TTAAAACTTTATGACTCCATAACTTTTAGACCGTGAGGCCAAACTTGGCA
Campell    TTAAAACTTTATGACTCCATAACTTTTAGACCGTGAG-CCAAACTTGGCA

chr3       TGTGATACCTTTTCGGAAAGCCCAAAGAGAGCTCTCCGATGAATGAGGAA
chr4       TGTGATACCTTTTCGGAAAGCCCAAAGAGAGCCCTCCGACGAAAGAAGCA
chr5       TGTGATACCTTTTCGGAAAGCCCAAAGACAGCCCTCCGACGAAAGAAGCA
Campell    TGTGATACCTTTTCGGAAAGCCCAAAGACAGCCCTCCGACGAAAGAAGCA

chr3       GGAAAATGGAATTCTTCTATTGTTTTTTTTCTACTCCAAATTTTGACCT
chr4       AGACAATGGAACTTTTCCATTGACTTTTTGTCGACTCCAAATTTTGACCT
chr5       GGACAATGGAACTTTTCCATTGACTTTTTGTCGACCCCAAATTTTGACCT
Campell    GGACAA-----CTTTTCCATTGACTTTTTGTCGACCCCAAATTTTGACCT

chr3       TAATGTACTTTTTCGGGCCTTTTCGTGATCTTTGCTATATTACGGGGCGA
chr4       TTAAGTACTTTTTCGGGCATTTTCGTGATTTGTGCTATATTACGGACCCA
chr5       TTAAGTACTTTTTCGGGCATTTTCGTGATTTGGGCTATATTACGGACCCA
Campell    TTAAGTACTTTTTCGGGCATTTTCGTGATTTGGGCTATATTACGGACCCA

chr3       AAATTATATGTTCGGACATTGTTTTCGAATATTTTGCATGTATCAAAGCT
chr4       AAATTACTTTTTCAAGCATTGTTTTCGAATATTTTTCATGCATCAAAGCT
chr5       AAATTACTTGTTCAAGCATTGTTTTCGAAT-TTTTTCATGCATCAAAGCT
Campell    AAATTACTTGTTCAAGCATTGTTTTCGAAT-TTTTTCATGCATCAAAGCT

chr3       CTTATAAACAAGATAGGGTATTCCTACATAGCTGGTGGGACCCACGGCAA
chr4       CGTTAAGACTAGATGGGGTATCCCTACATGGCGGGTGGGACCCACGGCGA
chr5       CGTTAAGACTAGATGGGGTATCCCTACATAGCGGGTGGGACCCACGGCGA
Campell    CGTTAAGACTAGATGGGGGATCCCTACATAGCGGGTGGGACCCACGGCGA

chr3       ATGGATCATAAAGTCTTCAAAAAAGAATAGATACGATTGCATTGCATGTA
chr4       ACGGTTCATCAAGACTTAAAAAAGAATAGATACGATTGCATTGCATATA
chr5       ATGGTTCATCAAGTCTTCAAAAAAGAATATATACGATTGCATTGCATATA
Campell    ATGGTTCATCAACTCTTCAAAAAAGAATATATACGATTGCATTGCATATA

chr3       GTAAC
chr4       CTAAC
chr5       CTAAC
Campell    CTAAC
```

Supplementary Figure 3

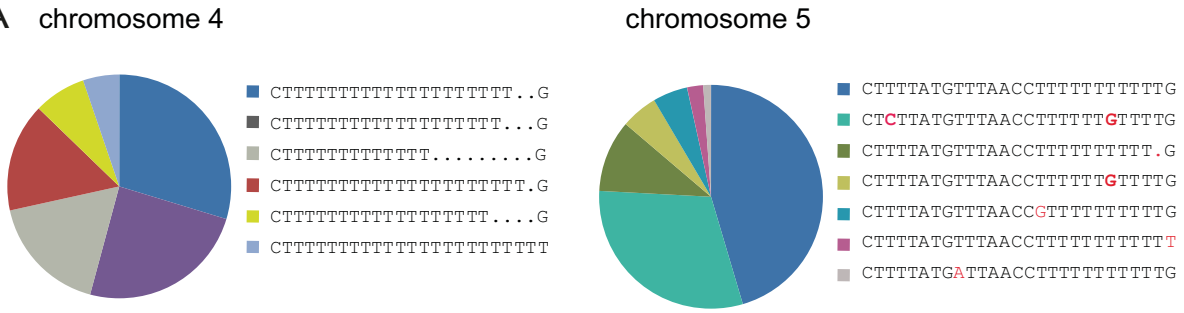


Supplementary Figure 4

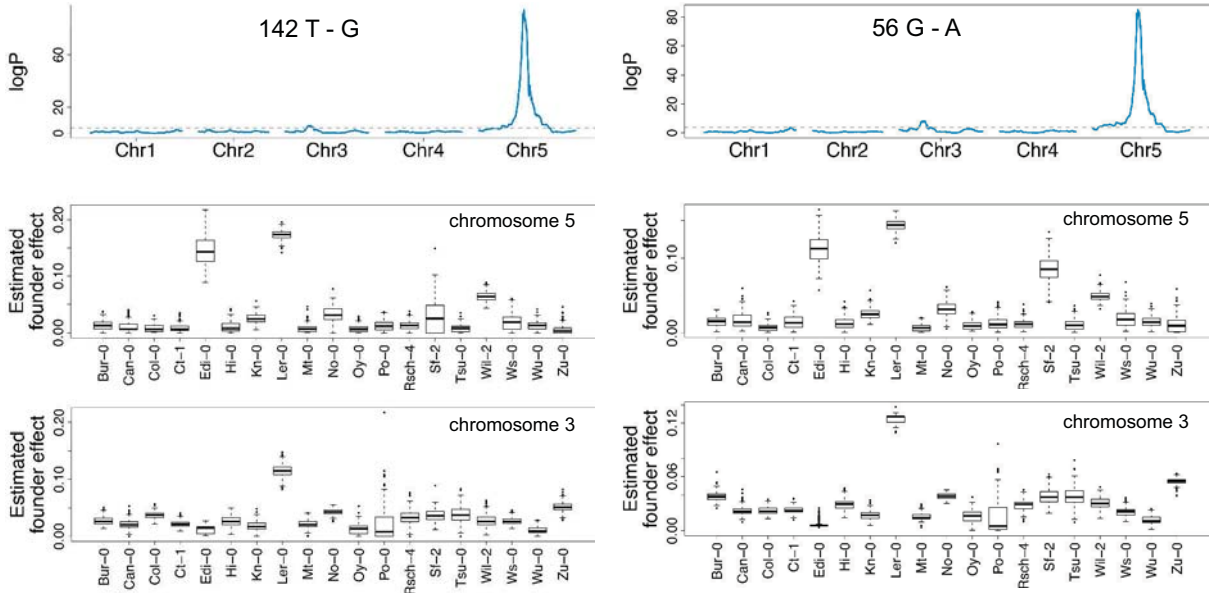


Supplementary Figure 5A-B

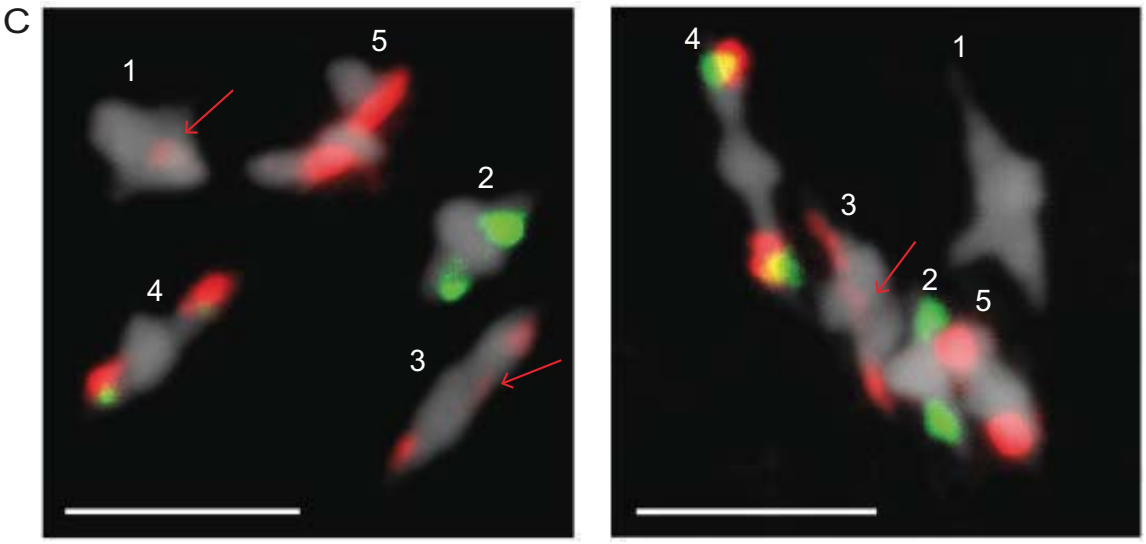
A

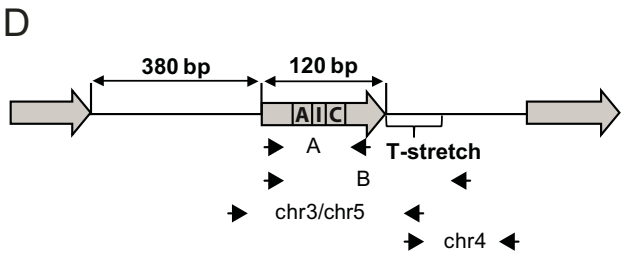


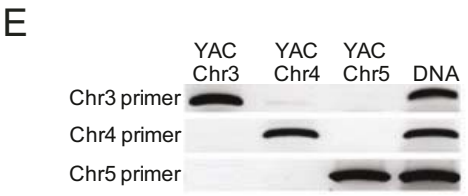
B

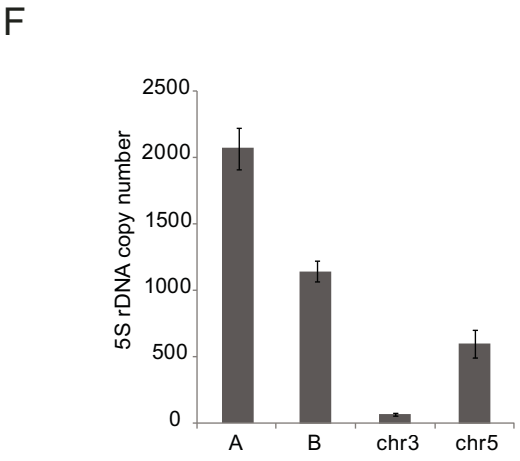


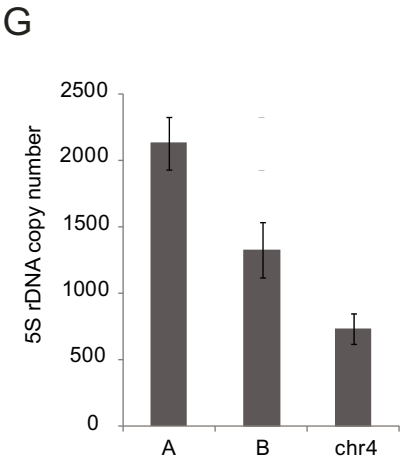
Supplementary Figure 5C-G



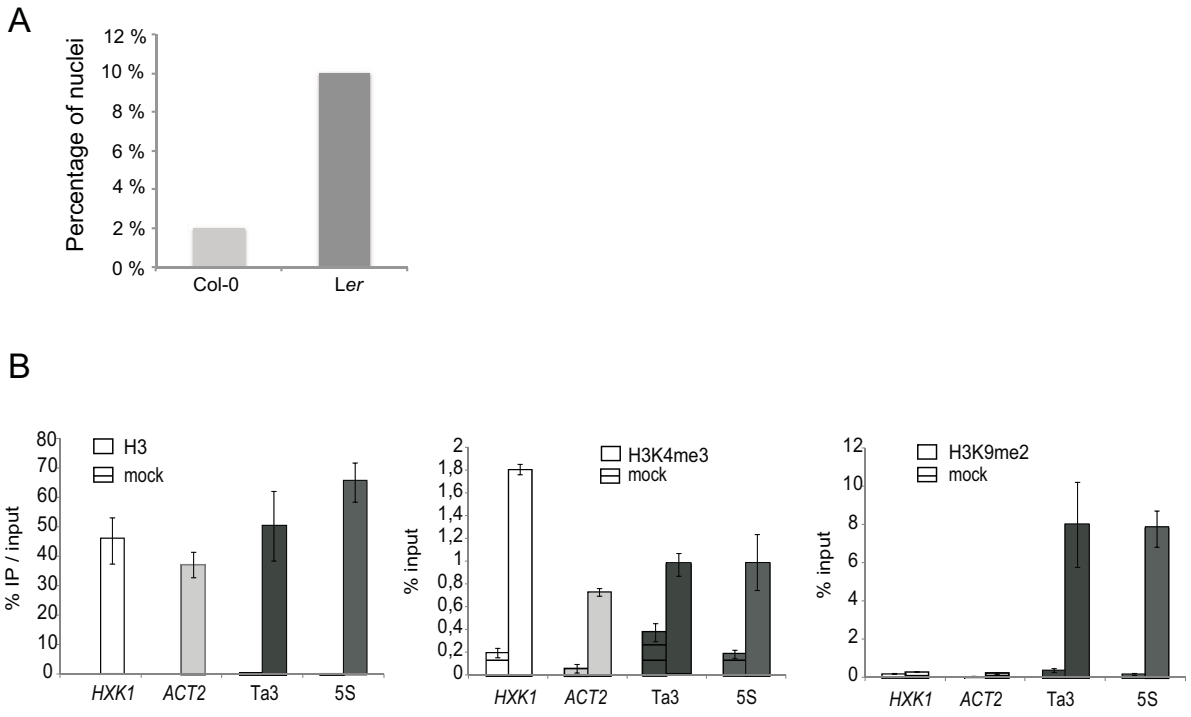








Supplementary Figure 6



Supplementary Table 1

Col-0	SRR1945757	The 1001 Genomes Consortium, 2016	Hi-seq
	to be deposited	this study	
	ERR420402 - ERR420412	Hagmann et al., 2015	

Ler	**Calculation of loci ratios**		
	ERR031544	Gan et al, 2011	Hi-seq
	SRR352145	Lu et al, 2011	
	DRR003391	Abe et al, 2015	
	Determination of polymorphisms		
	SRR3166543	Zapata et al, 2016	Hi-seq
	SRR3157034		
	Determination of T-stretchs		
	DRR003391	Abe et al, 2015	Hi-seq

Supplementary Table 2

ChIP-seq datasets

H3K27me1	SRR094101	Jacob et al, 2009
	SRR094102	
	SRR094103	
H3K36ac	SRR2932297	Mahrez et al, 2016
H3K4me3	SRR1964977	Brusslan et al, 2015
	SRR1964979	
H3K36me3	SRR3087130	Bewick et al, 2016
H3K56ac	SRR3087131	
H3K9me2	SRR3087128	
	SRR1999292	Stroud et al, 2014
H3	SRR1999291	
	SRR1005403	
Input H3.1	SRR394081	Stroud et al, 2012
Input H3.3	SRR394080	
H3.1	SRR394079	
H3.3	SRR394078	
Mnase	SRR957780	Li et al, 2014
H2A	SRR988543	Yelagandula et al, 2014
H2A.x	SRR988545	
H2A.Z	SRR988546	
H2A.W	SRR988544	
Input	SRR988541	

RNA-seq datasets

Shoot tissue from	SRR1030234	Zhang et al., 2015
	SRR1030235	

Supplementary Table 3

Primers used in this study

	Forward primer	Reverse primer
A	GGATGCGATCATACCAG	GGGAGGTCACCCATCCTAGT
B	GGATGCGATCATACCAG	CGAAAAGGTATCACATGCC
chr3	GACCCAYGGCGAATGGTT	ACCAAAAAAAAAAAAGGTAAACAT
chr4	AAAGTACWTWWAGGTCAAAATTTGGRG	TCCCTCTTTTTTTTTTTTTTTTTTTG
chr5	GGACCCAGGACAAATGGA	AARACAAAAAAAAAAGAGAAAAACCG
210	GGATGCGATCATACCAG	CGAAAAGGTATCACATGCC
HXK1	AGGAGCTCGTCTCTCTGCTG	GCTCAAACAATCCACCATCC
UEV1C	GGTGACTGAAATGTGAATTTGC	ATGCAGCCATCTCCTTCTTC

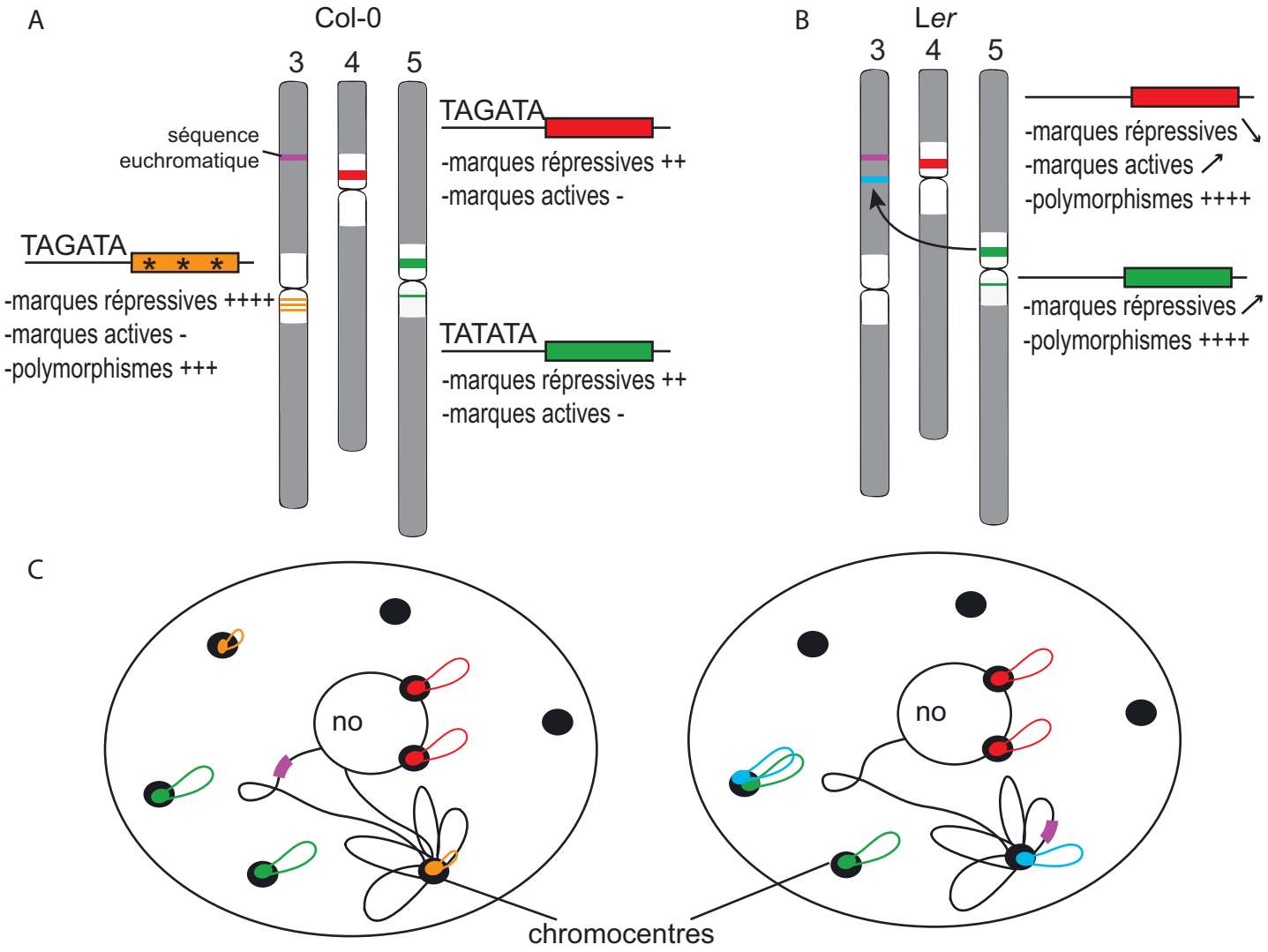


Figure 21 : Schéma récapitulatif de l'organisation, de la variabilité et du statut épigénétique des gènes d'ARNr 5S entre Col-0 et L*er*.

A) Les gènes d'ARNr 5S chez Col-0 sont globalement enrichis en marques épigénétiques répressives et dépourvus de marques permissives pour la transcription. Les gènes du chromosome 3 sont globalement plus enrichis en marques répressives par rapport aux loci du chromosome 4 et du chromosome 5. Les gènes d'ARNr 5S du chromosome 3 sont aussi les plus polymorphes. On trouve une mutation de la boîte TATA (TATATA en TAGATA) sur les gènes du chromosome 3 et du chromosome 4.

B) Les gènes d'ARNr 5S chez L*er* sont plus mutés que ceux de l'écotype Col-0 et les séquences portant le T-stretch du chromosome 4 sont dépourvues de marques répressives et enrichies en marques permissives par rapport à Col-0.

C) Dans l'écotype L*er*, les loci situés sur le bras du chromosome 3 ont tendance à s'associer aux chromocentres et aux ADNr 5S du chromosome 5.

Discussion et perspectives

Ma thèse s'est intéressée à l'organisation et la régulation de l'hétérochromatine et plus précisément de l'ADNr 5S chez *A. thaliana*. Mon travail s'est majoritairement concentré sur l'ADNr 5S, sur l'acquisition de connaissances sur son statut épigénétique, son organisation, et les types de séquences ADNr 5S qui existent dans les différents *loci* chez Arabidopsis.

Nos résultats (Figure 21) basés sur des analyses de données "whole genome", des analyses moléculaires et d'imagerie ont permis :

- de confirmer et d'affiner nos connaissances sur les gènes d'ARNr 5S en particulier concernant les polymorphismes des séquences transcrites et des T-stretchs entre les différents clusters d'ADNr 5S
- de déterminer le statut épigénétique de l'ADNr 5S en général et de manière spécifique pour chaque cluster.
- de montrer qu'il existe une grande diversité du nombre de séquences d'ADNr 5S entre différents écotypes d'*Arabidopsis* aussi bien dans le nombre global de copies que dans la proportion de copies d'ADNr 5S dans chaque cluster.
- de mettre en évidence des translocations de clusters d'ADNr 5S chez L*er* et certains mutants de la voie RdDM et de déterminer les origines des *loci* supplémentaires
- de montrer que les translocations affectent la localisation des *loci* 5S au sein du noyau en interphase
- de suggérer par des analyses de ChIP-qPCR une expression différentielle des différents *loci* chez Col-0 et L*er*

I. Polymorphismes des gènes d'ARNr 5S

L'organisation en répétitions en tandem des gènes d'ARNr 5S rend les études épigénomiques difficiles, notamment parce qu'il n'existe pas de séquence de référence pour les différents clusters. La majorité des données sur l'ADNr 5S ont été obtenues jusqu'à présent sur peu de séquences, issues de YAC et de BAC (Cloix et al., 2000). L'avantage des YAC et des BAC est qu'il est possible de les cartographier sur la carte génétique et physique du génome. Toutefois, ils présentent aussi plusieurs inconvénients : ils ne représentent

qu'une partie des clusters d'ADNr 5S, ils peuvent avoir subi des réarrangements et lorsqu'ils sont trop riches en séquences répétées il est souvent difficile de les inclure dans l'assemblage du génome. Il était donc important d'apporter des informations complémentaires sur les clusters d'ADNr 5S grâce aux nouveaux moyens technologiques à notre disposition. La première chose que nous avons vérifiée a été les séquences des T-stretchs (**Annexe III**). Nous avons validé les séquences identifiées par Cloix. Cependant, pour les T-stretchs des trois *loci* du chromosome 3, bien que la signature du *locus* 1 et 3 du chromosome 3 soit la plus représentée des signatures trouvées dans notre étude, la signature du *locus* 2 du chromosome 3 ne représente que 2% des séquences du chromosome 3 (c'est le 12ème T-stretch en abondance). La question reste donc ouverte sur l'existence de 3 *loci* d'ADNr 5S sur le chromosome 3. En effet, la pseudomolécule TAIR10 présente 6 *loci* sur ce chromosome, mais étant donné l'assemblage incomplet des gènes d'ARNr 5S sur la pseudomolécule, il est possible qu'il existe bien 3 *loci* ou bien qu'il n'existe qu'un seul cluster présentant des unités avec des T-stretch mutés comme c'est le cas pour les autres chromosomes.

Nous avons pu observer grâce aux sondes LNA/DNA mixmers que les séquences ADN avec le même T-stretch sont majoritairement localisées au niveau du même cluster ce qui confirme que les T-stretchs sont enrichis de manière chromosome spécifique chez *A. thaliana*. Nous avons aussi montré que les T-stretchs du chromosome 3 sont les plus polymorphes des trois *loci*. Nous avons observé une conservation des nucléotides spécifiques des T-stretchs entre les différents écotypes d'*A. thaliana* analysés. A l'inverse, l'espèce proche *Arabis alpina* possède un T-stretch différent.

La spécificité des T-stretchs nous a permis par la suite d'analyser spécifiquement les polymorphismes des gènes d'ARNr 5S et leur statut épigénétique pour chaque *locus*. Cette analyse (**Annexe IV**) a montré que les séquences portant la signature du chromosome 3 sont les plus polymorphes ce qui est en accord avec les résultats antérieurs (Cloix et al., 2003). Cette analyse a aussi confirmé la grande conservation des séquences transcrites et le plus faible taux de polymorphismes présent sur les *loci* des chromosomes 4 et 5 chez Col-0 ce qui explique en partie l'absence de séquence de référence pour les clusters d'ADNr 5S pour ces deux clusters dans la pseudomolécule TAIR10. Malgré cette grande conservation, la présence de T-stretchs et de polymorphismes spécifiques de chaque *locus* est en accord avec des études précédentes suggérant qu'un équilibre entre la conversion de gènes et une pression de sélection contre des gènes polymorphes permet une forte homogénéité des

séquences au sein d'un *locus* et une variabilité entre les *loci* (Eickbush and Eickbush, 2007).

Plus précisément, lorsqu'une mutation apparait au sein d'un *locus* d'ADNr 5S, celle-ci pourrait soit être étendue par conversion génique au sein d'un *locus* soit au contraire être éliminée par la pression de sélection si elle est délétère (Eickbush and Eickbush, 2007).

En utilisant deux jeux de données indépendants de séquençage de L*er* (Zapata et al., 2016), nous avons aussi pu montrer que la séquence transcrite des gènes d'ARNr 5S présente plus de polymorphismes dans l'écotype L*er* que dans Col-0. Les polymorphismes se trouvent principalement en 3' de la séquence transcrite avec notamment la présence d'insertions. Nous avons pu confirmer certains de ces polymorphismes grâce à l'analyse des 'MAGIC lines' en collaboration avec F. Rabanal.

L'analyse des séquences transcrites présentes dans l'écotype L*er* a dévoilé beaucoup de polymorphismes comparés aux séquences consensus du chromosome 4 et du chromosome 5. Plus de 95 % des séquences analysées chez L*er* possèdent plus de 2 polymorphismes comparés aux séquences consensus de Col-0. Certains de ces polymorphismes sont présents au niveau des hélices de l'ARNr 5S ou encore en 3' de la séquence transcrite (notamment des insertions) ce qui pourrait modifier la structure de l'ARNr 5S ou encore sa longueur. Nous ne savons pas, dans l'état actuel de nos connaissances si ces polymorphismes ont un impact sur la transcription des gènes d'ARNr 5S ou sur la conformation et la fonction de l'ARN 5S dans le ribosome.

Afin de savoir quels gènes d'ARNr 5S sont transcrits dans les écotypes Col-0 et L*er* ainsi que la contribution respective de chaque cluster dans la production des ARNr 5S de la cellule, nous sommes en train de réaliser un séquençage des ARNr 5S et des transcrits de 5S-210 par RNA-Seq. A partir d'un tel séquençage à forte profondeur, nous pourrons grâce à notre étude de polymorphismes déterminer l'origine des ARNr 5S transcrits (entre clusters et au sein d'un cluster) ainsi que la proportion d'ARNr 5S majeurs/mineurs produits dans chaque écotype.

Parmi les polymorphismes identifiés, nous avons détecté des mutations au niveau de la boîte TATA-like (TAGATA plutôt que TATATA) sur les séquences du chromosome 3 et celles du chromosome 4 (respectivement 82% et 56.8% des séquences analysées). Cette mutation avait déjà été identifiée sur le chromosome 3 (Cloix et al., 2003) et est connue pour diminuer l'expression des gènes d'ARNr 5S *in vitro* de 54% (Cloix et al., 2003). Il serait intéressant de regarder si les gènes présentant une boîte TATA-like mutée sur le

chromosome 4 sont des séquences majoritaires ou au contraire des séquences associées à des polymorphismes. L'absence de cette mutation sur le cluster du chromosome 5 pourrait suggérer une activité transcriptionnelle plus importante par rapport au chromosome 4 ce qui serait en accord avec le niveau des marques permissives pour la transcription légèrement plus élevé sur les gènes du *locus* du chromosome 5 par rapport au celui du chromosome 4. Une autre possibilité afin de déterminer si l'un des *loci* est plus efficacement transcrit, serait d'utiliser un anticorps qui reconnait le dimère ADN/ARN. En effet, un tel anticorps est capable de différencier les *loci* d'ADNr 45S qui sont transcrits de ceux qui sont silencieux de par la très forte transcription de ces gènes et donc de la forte présence de dimères ADN/ARN sur ces *loci* (Peter Schloegelhofer, *communication personnelle*). Il serait possible d'utiliser cet anticorps couplé aux sondes LNA/DNA mixmers *locus* spécifiques afin de savoir si le *locus* du chromosome 4 et/ou celui du chromosome 5 est/sont transcrits. D'après le séquençage de transcrits de 5S-210 dans l'écotype Col-0 (Douet et al., 2009), les transcrits 5S-210 proviennent majoritairement du *locus* du chromosome 5. Il serait donc tentant de spéculer que le *locus* du chromosome 5 soit plus activement transcrit que celui du chromosome 4 peut-être du fait d'un état chromatinien plus "relâché". Une « dominance transcriptionnelle » d'un *locus* sur un autre a été observée pour l'ADNr 45S où seul le *locus* du chromosome 4 est transcrit et forme le nucléole et pas celui du chromosome 2 chez Col-0 (Chandrasekhara et al., 2016). De façon intéressante, dans un mutant pour la chaperonne d'histone CAF-1, les variants 1 (VAR1) d'ADNr 45S sont exprimés et le cluster du chromosome 2 (région NOR2) est retrouvé proche du nucléole contrairement à une plante sauvage où le VAR1 est réprimé et le NOR2 n'est pas associé au nucléole (Pontvianne et al., 2013). Il semble donc que pour les *loci* d'ADNr 45S le *locus* du chromosome 2 soit potentiellement "activable" mais maintenu sous silence en condition normale. Un scenario similaire pourrait être suggéré pour les *loci* d'ADNr 5S. Afin d'étudier cette hypothèse, il conviendrait de comparer Col-0 aux autres écotypes ou d'étudier Col-0 dans différents mutants altérant la structure chromatinienne.

À ce point de notre étude, au niveau génomique, il reste à déterminer si toutes les séquences d'un *locus* présentent le même T-stretch, même si ceci est fortement suggéré par les résultats des analyses FISH. Notre objectif final reste d'obtenir une séquence de référence complète des *loci* d'ADNr 5S. Cet objectif reste toutefois très difficile à atteindre à cause de limites techniques, notamment l'impossibilité pour les outils d'assemblage *de novo* disponibles (Abyss, SOAPdenovo, Velvet…etc) d'assembler des séquences répétées hautement conservées. Notons cependant que la société NRgene a développé un nouvel

outil qui, à partir de données Hi-Seq, a réussi à assembler plus de 120 génomes en 2016 notamment celui du blé particulièrement complexe (http://nrgene.com). Cet outil très prometteur pourrait enfin permettre l'assemblage plus complet du génome d'*A. thaliana*. Nous pensons que l'évolution des techniques de séquençage avec la création de reads de plus en plus longs (PacBio, Nanopore) (Clarke et al., 2009; Eid et al., 2009) et de moins en moins 'error-prone' ainsi que l'évolution des logiciels d'assemblage permettra peut-être à l'avenir d'obtenir une séquence du génome complètement assemblée et ce même au niveau des séquences répétées du génome.

II. Variation du nombre de copies

À propos des difficultés techniques

Lorsque nous avons mis en place les outils bioinformatiques dédiés à l'analyse des gènes d'ARNr 5S, nous avons été confrontés à des difficultés méthodologiques et techniques. L'utilisation de différents logiciels de mapping (bowtie, BWA-ALN, BWA-MEM) nous a montré qu'en fonction de l'algorithme utilisé, les résultats obtenus pour les valeurs absolues du nombre de copies d'ADNr 5S pouvaient être différents. C'est par exemple le cas des 2 versions de BWA (BWA-MEM et BWA-ALN) qui utilisent des algorithmes très différents.

Lorsque nous avons utilisé la séquence transcrite (120pb) comme séquence de référence, nous avons rencontré un problème de normalisation avec BWA-ALN puisque les reads plus longs que la séquence de référence sont éliminés (score plus faible). Pour exploiter notre séquençage Mi-Seq il fallait donc découper nos reads avant l'analyse par BWA-ALN, ce qui n'est pas satisfaisant. Lorsque nous avons voulu utiliser BWA-MEM qui est adapté aux reads plus longs, nous avons été confrontés à un autre problème qui était l'impossibilité de définir le nombre de mismatchs autorisés. En effet, BWA-MEM calcule des scores d'alignement et ne sélectionne que les meilleurs, ce qui se traduit par une perte des séquences hautement polymorphes. Ces problèmes techniques très spécifiques et relatifs au fonctionnement des algorithmes ont montré les limites de nos compétences en bioinformatique et nécessiteraient une collaboration plus étroite avec des spécialistes des outils de mapping afin de prendre en compte nos spécificités (séquences de références courtes, peu de polymorphisme, répétitions importantes du nombre de copie/génome).

Malgré le fait que nous disposions de reads Hi-Seq (50pb) et Mi-Seq (300pb), ces difficultés nous ont conduit à restreindre notre analyse à l'utilisation des reads de 50pb notamment en découpant les reads plus longs dans certains cas. Nous avons préféré en effet, conserver le même algorithme (BWA-MEM) pour exploiter les différentes données disponibles pour Col-0, L*er* ou l'ensemble des écotypes des « 1001 genomes ».

Nous avons aussi observé que la profondeur de séquençage influençait les résultats, avec de meilleurs résultats et reproductivité entre des data sets ayant une grande profondeur de séquençage.

Nous nous sommes aussi aperçus qu'il n'était pas possible de réaliser certaines étapes de nettoyage normalement réalisées en routine lors de l'analyse de données de séquençage. En effet, l'utilisation d'outils de nettoyage des duplicats de PCR n'est pas possible pour l'ADNr 5S puisque la répétition de ces séquences entraine la perte de la quasi totalité des reads d'ADNr 5S. Il n'est pas non plus possible de filtrer les reads avec un niveau de qualité (QC) trop élevé car la répétition des thymines au niveau du T-stretch semble faire baisser grandement la qualité des reads d'ADNr 5S et plus particulièrement ceux du chromosome 4 qui présente une suite ininterrompue de thymines.

En résumé, afin d'obtenir une estimation du nombre de copies la plus exacte et comparable entre toutes les données de séquençage disponibles, nous avons opté pour la méthode BWA-MEM, utilisant des reads de 50 bp avec une normalisation sur un set de 18 gènes monocopies (Czechowski et al., 2005).

À propos de nos résultats

En 1992, Campell a estimé le nombre de gènes d'ARNr 5S chez *A. thaliana* à 1000 copies par génome haploïde. Cette estimation a été faite par Southern blot et avec une estimation de la taille du génome haploïde à 70 Mb alors que l'estimation actuelle serait de l'ordre de 125 Mb aujourd'hui (Leutwiler et al., 1984, *https://www.arabidopsis.org/*). Nos analyses *in silico* et moléculaires ont révélé l'existence de plus de 2000 copies d'ADNr 5S dans l'écotype Col-0 ce qui est largement supérieur au nombre estimé par Campell. D'après notre estimation, les gènes d'ARNr 5S représentent 0.7% du génome de l'écotype Col-0.

Nous avons aussi observé une très grande variation du nombre de copies d'ADNr 5S entre différents écotypes de l'espèce *A. thaliana*, que ce soit au niveau du nombre de copies totales (de 241 à 15 738 copies) ou au niveau du nombre de copies par cluster. Il est

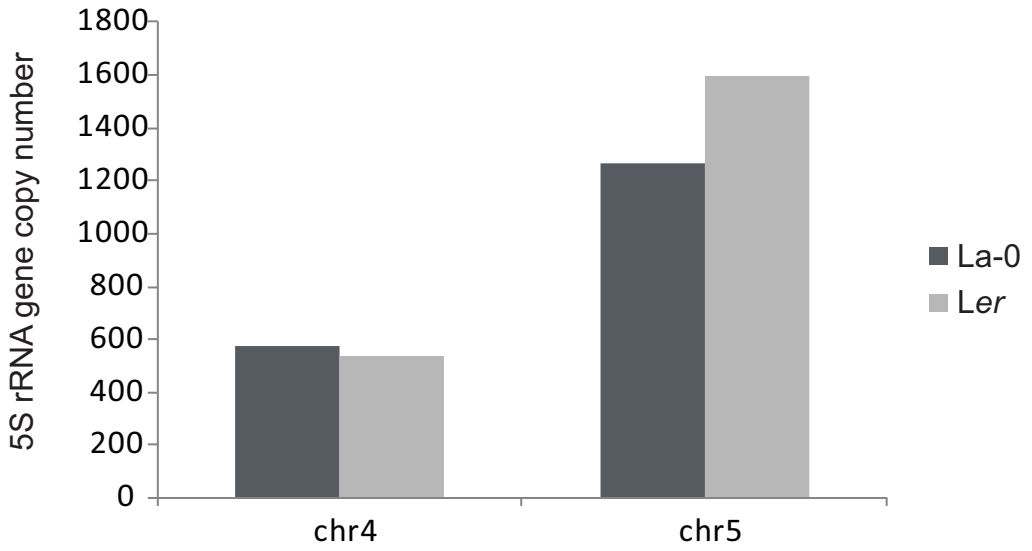
possible qu'il existe un nombre minimum d'ADNr 5S nécessaire pour la survie de la plante et que, dans les écotypes avec très peu d'ADNr 5S, seuls les gènes majoritaires non mutés aient été conservés.

Nous n'avons observé qu'une très faible corrélation entre le nombre de copies d'ADNr 5S et 45S dans les écotypes étudiés. Il semble donc que les régulations du nombre de copies d'ADNr 5S et 45S soient indépendantes l'uns de l'autre.

Si les *loci* des chromosomes 4 et 5 semblent bien conservés, seulement 8 des 34 écotypes étudiés présentent des séquences avec la signature du chromosome 3. Ce cluster est petit comparé aux autres *loci*, il n'est pas transcrit (Cloix et al., 2002) et est fortement polymorphe dans l'écotype Col-0. Ce *locus* est-il nécessaire et participe-t-il à une fonction particulière par exemple dans l'organisation et la structuration du génome au sein du noyau ? Deux hypothèses peuvent être formulées : soit il est apparu récemment par translocation soit il tend à disparaitre à partir d'un évènement de duplication plus ancestral. Afin de trancher entre ces deux hypothèses, il faudrait étudier plus attentivement l'histoire évolutive des 34 écotypes étudiés. Il serait également intéressant de reprendre les données de duplications du génome d'*A. thaliana* afin de savoir si le cluster du chromosome 3 appartient à un segment dupliqué au sein duquel les régions dupliquées tendent à être éliminées afin de revenir vers un état diploïde. Ce phénomène classique dans l'évolution des génomes montre que de telles régions ont une vitesse évolutive plus élevée, ce qui semble être le cas pour le cluster du chromosome 3 qui a accumulé de nombreuses mutations. Le chromosome 3 semble donc échapper au processus d'homogénéisation par conversion génique observé pour les chromosomes 4 et 5.

Nous avons aussi pu observer des différences dans le nombre et la localisation des copies d'ADNr 5S dans des mutants de la voie RdDM et dans l'écotype L*er.* Les mutants pour les protéines AGO4, DICER-LIKE2 et DICER-LIKE3 présentent un gain ou une perte de *loci* d'ADNr 5S. Ces protéines sont impliquées dans la mise en place de la méthylation de l'ADN et par boucle d'amplification avec les protéines SUVH, dans la mise en place de la marque H3K9me2. Ces mutants qui présentent une chromatine moins compacte qu'une plante sauvage pourraient favoriser les évènements de recombinaison, on sait notamment que la méthylation de novo par la voie RdDM est suffisante pour maintenir sous silence des "hot-spots" de recombinaison (Yelina et al., 2015).

A



B

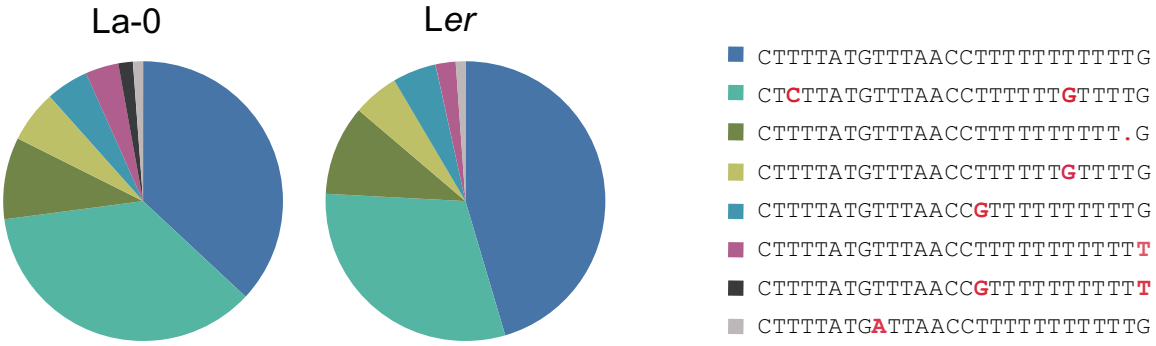


Figure 22 : Nombre de copies et polymorphismes des ADNr 5S portant la signature du chromosome 5 entre l'écotype L*er* et La-0

A) Nombre de copies d'ADNr 5S chez La-0 et L*er*.

B) Polymorphismes retrouvés au niveau des T-stretchs du chromosome 5 chez La-0 et L*er*.

Dans les mutants *ago4* analysés ainsi que dans l'écotype L*er*, l'apparition ou la disparition d'un *locus* d'ADNr 5S concerne les séquences portant le T-stretch du chromosome 5. Il est possible que ces séquences soient les plus touchées de par leur état chromatinien et/ou transcriptionnel. En effet, le cluster du chromosome 5 n'a pas de polymorphisme dans la boite TATA-like et est enrichi en marques H3K4me3, H3K56ac et H3K36me3. Ces différences en comparaison du cluster du chromosome 4 laisse penser qu'une grande partie des gènes d'ARNr 5S du type majoritaire sont transcrits.

En étudiant plus particulièrement la présence d'un *locus* d'ADNr 5S sur le bras chromosomique du chromosome 3 dans l'écotype L*er*, nous avons pu montrer que ces séquences 5S portent la signature du *locus* du chromosome 5. Il s'agirait donc d'une translocation du chromosome 5 vers le chromosome 3. L'analyse des polymorphismes nous a permis de montrer qu'il existe des T-stretchs spécifiques de l'écotype L*er* par rapport à Col-0 et ce polymorphisme a été mis à profit pour cartographier ce nouveau *locus* à l'aide des 'MAGIC lines'. Le nouveau *locus* est localisé près d'une inversion de 170kb chez L*er*. L'historique de l'écotype Ler montre qu'il dérive de Landsberg (La) après irradiation (Zapata et al., 2016). Il est probable que cette inversion ait été générée lors de l'irradiation et que ce réarrangement génomique ait entrainé une duplication ou une translocation d'un segment du chromosome 5 portant une partie du cluster d'ADNr 5S du chromosome 5. Une autre hypothèse impliquerait des éléments transposables dans la translocation des gènes d'ARNr 5S. En effet, il existe dans le génome un rétrotransposon appelé Cassandra comportant une portion des gènes d'ARNr 5S (Garcia et al., 2010; Richard et al., 2008). Cependant ce rétrotransposon ne comportent pas la séquence T-stretch et il est donc peu probable qu'il soit impliqué dans le déplacement spécifique d'un *locus* 5S vers le chromosome 3 ou ailleurs dans le génome. De plus, des translocations d'ADNr 5S ont été observées au sein de certains espèces qui ne comportent pas le rétrotransposon Cassandra (Wicke et al., 2011).

Nous avons trouvé plus de copies d'ADNr 5S portant le T-stretch du chromosome 5 par analyse bio-informatique dans L*er* que dans l'écotype Landsberg ce qui suggérerait plutôt une duplication d'une partie du cluster du chromosome 5 sur le chromosome 3 (Figure 22). Concernant les T-stretchs, il y a peu de différences entre La-0 et L*er* ce qui ne nous a pas permis de conclure si la partie dupliquée était enrichie en séquences portant un T-stretch précis.

III. Régulation épigénétique des gènes d'Arabidopsis thaliana

Afin de produire la quantité nécessaire d'ARNr 5S, les gènes d'ARNr doivent être finement régulés notamment les gènes minoritaires qui ne sont exprimés qu'à certains moments du développement de la plante (Layat et al., 2012a; Mathieu et al., 2003).

Notre étude a montré que les gènes d'ARNr 5S dans l'écotype Col-0 sont enrichis en nucléosomes et en marques épigénétiques répressives pour la transcription. Les gènes présents sur le chromosome 3 sont globalement plus enrichis en marques répressives que les autres *loci* ce qui est en accord avec le fait qu'ils soient en position péricentromérique et qu'ils ne soient pas transcrits. De plus, nous avons pu observer que les gènes d'ARNr 5S du chromosome 4 et du chromosome 5 sont enrichis en marque H3K27me1 comparés aux gènes d'ARNr 5S du chromosome 3 qui sont enrichis en marque H3K9me2. H3K9me2 contribuerait à l'établissement d'un environnent transcriptionnellement répressif sur le chromosome 3 alors que la marque H3K27me1 permettrait la régulation de la transcription des *loci* des chromosomes 4 et 5. La régulation des *loci* 5S serait donc variable entre les différents clusters/chromosomes.

Les profils épigénétiques du *locus* du chromosome 4 et du chromosome 5 sont similaires et ne permettent pas de savoir si un *locus* est plus favorable à la transcription que l'autre même si les marques H3K4me3, H3K56ac et H3K36me3 sont plus enrichies sur le chromosome 5 par rapport au 4. Bien que la littérature ne suggère pas l'existence de marques épigénétiques agissant spécifiquement sur la transcription par l'ARN polymérase II et par l'ARN polymérase III (Barski et al., 2010; Bhargava, 2013; White, 2011), nous ne pouvons pas exclure à ce stade de notre étude que certaines marques n'aient pas une activité spécifique sur la transcription des gènes d'ARNr 5S ou qu'il existe des marques épigénétiques spécifiques de l'ARN polymérase III comme c'est le cas chez la levure pour la marque H2AQ105me qui est enrichie exclusivement au niveau des gènes d'ARNr 35S (Tessarz et al., 2014).

Si nous avons actuellement une meilleure vision de l'enrichissement en marques épigénétiques entre les clusters d'ADNr 5S, il reste à étudier plus finement la régulation au sein d'un cluster. Si dans le futur des analyses de ChIP-seq générant des reads longs deviennent disponibles, il sera intéressant d'étudier l'enrichissement différentiel en marques épigénétiques des gènes majoritaires, minoritaires et hautement polymorphes le long d'un

cluster. Ceci pourrait nous permettre également de confirmer que les séquences présentes sur les boucles d'ADNr 5S, détectées par FISH dans les noyaux en interphase, sont les séquences majoritaires puisqu'il a été montré par immunofluorescence qu'elles sont enrichies en marques H3K9ac et H3K4me2.

IV. L'action des chaperonnes d'histone

Dans l'article en **annexe VI.A** de cette thèse, nous avons identifié les orthologues des protéines du complexe HIR et étudié leur rôle dans l'intégrité du génome. Les plantes mutantes pour la protéine HIRA sont les plus affectées phénotypiquement laissant penser que HIRA est la protéine centrale du complexe HIR. Le mutant *hira* présente une baisse dans l'occupation nucléosomale dans le corps et la région 3' des gènes et des séquences hétérochromatiques ce qui confirme son rôle de chaperonne. Notre analyse a permis de montrer que HIRA contribue au maintien sous silence des séquences hétérochromatiques et ce malgré le fait que HIRA soit une chaperonne de l'histones H3.3. Il est possible que HIRA soit responsable du dépôt d'histone H3.3 sur les séquences hétérochromatiques afin de garder une occupation nucléosomale correcte et permettant d'assurer une régulation de la transcription.

En effet, une baisse de l'occupation nucléosomale a été observée au niveau des gènes d'ARNr 5S dans le mutant *hira* mais celle-ci n'est pas associée à une activation de la transcription des ARN 5S-210 observée dans d'autres mutants épigénétiques (**Annexe VI.A**). Le variant d'histones H3.3 étant localisé au niveau des gènes transcrits, il est envisageable qu'il soit associé surtout aux gènes majoritaires. Par la suite il serait intéressant d'analyser si la perte de ce variant impacte la transcription des gènes majoritaires et entraîne une baisse globale du niveau d'ARNr 5S dans la cellule.

Nous avons aussi montré que la chaperonne HIRA ne semble pas impliquée dans la formation des chromocentres lors du développement de la plantule. En effet, la formation des chromocentres entre 2 et 5 jours du développement post-germination ne semble pas altérée par la mutation *hira* alors que c'est le cas dans des mutants du complexe CAF-1 (**Annexe V**).

Nous proposons que, en plus d'HIRA, la chaperonne d'histone ATRX soit impliquée dans le dépôt du variant d'histones H3.3 (**Annexe VI.A** et **VI.B**). Les plantes mutantes pour ATRX présentent une infertilité modérée, une baisse de l'occupation nucléosomale sur

certains gènes, notamment l'ADNr 45S, ainsi qu'une altération d'expression des variants 45S. Cependant, nous n'avons observé que peu de changements dans les mutants *atrx* au niveau de l'occupation nucléosomale sur les gènes d'ARNr 5S comme sur d'autres séquences centromériques ou péricentromériques. ATRX ne semble donc pas être une chaperonne d'histone impliquée dans la régulation de l'occupation nucléosomale de l'ADNr 5S. Par contre, son rôle dans la régulation de l'ADNr 45S pourrait signifier que différents mécanismes sont impliqués dans la régulation chromatinienne des gènes d'ARNr 5S et 45S. En effet, la perte de la chaperonne CAF-1 affecte exclusivement les gènes d'ARN 45S (Mozgová et al., 2010). Cette différence de régulation entre les gènes d'ARNr est cohérente avec nos observations sur le nombre de copies d'ARNr 5S et 45S qui ne semble pas être co-régulé chez *A. thaliana*.

V. Organisation nucléaire des gènes d'ARNr 5S au cours du développement

Les chromocentres se structurent entre 2 et 5 jours après germination dans les cotylédons (Benoit et al., 2013; Douet et al., 2008; Mathieu et al., 2003). Pendant cette fenêtre de temps, les gènes d'ARNr 5S subissent également une forte réorganisation. Actuellement, les mécanismes impliqués dans l'organisation des chromocentres et la régulation de cette réorganisation restent largement inconnus. Dans l'article Benoit et al. (**Annexe V**), nous suggérons que le dépôt des variants H3.1 au niveau des séquences répétées par le complexe CAF-1 serait un des éléments impliqués dans la formation des chromocentres. Nous avons aussi découvert que cette organisation nécessite la mise en place des marques épigénétiques H3K9me2 et H3K27me1. Ayant montré, par nos analyses de ChIP-seq, que ces marques épigénétiques sont enrichies au niveau des gènes d'ARNr 5S, il serait intéressant de regarder la dynamique de l'ADNr 5S dans un mutant *caf-1* ou dans des mutants pour les protéines impliquées dans le dépôt des marques épigénétiques comme ATXR5/ATXR6 et SUVH4, 5 et 6 ainsi que la proportion d'ARNr 5S minoritaires. Ceci pourrait nous aider à comprendre si la transcription des gènes d'ARNr 5S est régulée par ces différentes marques épigénétiques au cours du développement.

Nous disposons au laboratoire d'outils génétiques intéressants afin de répondre à ces questions : de nombreux mutants de chaperonnes d'histones, des protéines responsables de la mise en place de différentes marques épigénétiques ainsi que de plantes transformées avec des versions étiquetées de H3.1 et H3.3.

VI. Organisation des gènes d'ARNr 5S dans le noyau

Les observations réalisées avec nos sondes ‘LNA mixmers’ chromosome-spécifiques nous ont permis de montrer que le cluster présent sur le bras du chromosome 3 dans L*er* est majoritairement associé aux chromocentres et ce malgré sa localisation sur le bras chromosomique qui correspond à de l'euchromatine. Il est possible que les gènes d'ARNr 5S présents sur le bras long du chromosome 3 soient étiquetés épigénétiquement afin de leur conférer un statut "hétérochromatique" par la cellule et que ces marques contribuent à un recrutement du cluster du chromosome 3 vers les chromocentres. C'est le cas pour le knob du chromosome 4, qui est une séquence hétérochromatique sur le bras court du chromosome 4 provenant d'une inversion (Fransz et al., 2000, 2016). En effet le knob, bien que présent sur le bras du chromosome 4, a gardé ses marques épigénétiques hétérochromatiques et est associé aux régions péricentromèriques du chromosome 4 dans les noyaux en interphase (Fransz et al., 2016; Grob et al., 2013). Une étude chez l'homme a aussi montré que l'insertion d'un transgène portant un gène d'ARNr 5S a tendance à se localiser au niveau de la périphérie nucléaire et à diminuer l'expression d'un gène rapporteur attaché à ce gène d'ARNr 5S (Fedoriw et al., 2012).

Il est envisageable que les gènes proches du cluster d'ADNr 5S sur le bras long du chromosome 3 voient leur niveau de transcription altéré par la localisation du cluster au niveau des chromocentres. Néanmoins, le contraire a été observé pour les gènes au voisinage du knob pour lesquels les marques épigénétiques et le niveau de transcription ne sont pas impactés par la présence du knob. Afin de vérifier l'état transcriptionnel des gènes présents autour du *locus* d'ARNr 5S sur le chromosome 3, il faudrait affiner la cartographie du cluster d'ADNr 5S et comparer des données de RNA-seq entre Col-0 et L*er*.

Le *locus* d'ADNr 5S sur le chromosome 3 dans L*er* étant localisé au niveau du bras chromosomique, nous pensions observer une perte des marques épigénétiques répressives au niveau des gènes d'ARNr 5S portant le T-stretch du chromosome 5. A l'inverse, nous avons observé une perte de marques répressives et une augmentation des marques permissives pour la transcription au niveau des séquences du chromosome 4. Avec ces résultats, nous pouvons postuler que ce sont donc les gènes d’ARNr 5S du chromosome 4 qui sont transcrites principalement dans L*er.* Ceci pourrait s'expliquer par le fait que les séquences présentant le T-stretch du chromosome 5 sont plus polymorphes que les séquences portant les signatures du chromosome 4 chez L*er*, ou encore par le fait qu'il est

possible que la translocation des séquences du chromosome 5 vers le chromosome 3 ait créé des réarrangements au niveau du cluster du chromosome 5 entrainant une mise sous silence de ces séquences. Inversement, le fait que 50% de gènes d'ARNr 5S sur le chromosome 4 chez Col-0 portent une mutation dans la TATA box pourrait indiquer une transcription prédominante des gènes du chromosome 5 chez cet écotype. Il existerait donc des différences de régulation entre écotypes avec la dominance d'un *locus* sur un autre entre les clusters des chromosomes 4 et 5. L'étude de croisements entre écotypes comme L*er* et Col-0 pourrait donc s'avérer un outil intéressant pour étudier la mise en place de ce mécanisme. Pour les ADNr 45S, il existe un phénomène appelé "Nucleolar dominance" dans les plantes hybrides. Ce phénomène correspond à la formation du nucléole à partir d'un seul chromosome parental et pas à partir des 2 chromosomes (Pontes et al., 2003). Il a été montré par ailleurs qu'il s'agissait plutôt d'une mise sous silence épigénétique d'un chromosome (déacétylation, méthylation ADN) qu'une activation transcriptionnelle de l'autre chromosome (Lawrence et al., 2004; Preuss and Pikaard, 2007; Probst et al., 2004).

De nombreuses données restent à découvrir sur cette séquence modèle comme l'organisation des gènes d'ARNr 5S et leur potentiel enrichissement différentiel en marques épigénétiques le long d'un cluster, la régulation du nombre de copies d'ADNr 5S (qui ne semble pas corrélée aux nombres de copies d'ADNr 45S) ou encore les conséquences d'un point de vue transcriptionnel d'un changement du nombre de copies ou de la localisation des clusters d'ADNr 5S. D'autres facteurs régulateurs n'ont malheureusement pas pu être étudiés comme le rôle régulateur des petits ARNs, la fixation du facteur TFIIIA au niveau des différents clusters et nos données sur la transcription différentielle entre les différents *loci* n'est pas encore finalisée. L'ADNr 5S est une séquence dont l'étude est complexe de par son organisation en séquences répétées en tandem et sa localisation en région péricentromérique. Notre travail a permis de mieux cerner les difficultés d'une telle étude et, nous l'espérons, d'approfondir nos connaissances sur cette séquence ADN et ses mécanismes de régulation chez *Arabidopsis*.

Références

Abou-Ellail, M., Cooke, R., and Sáez-Vásquez, J. (2011). Variations in a team: major and minor variants of Arabidopsis thaliana rDNA genes. Nucleus **2**: 294–9.

AGI (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408**: 796–815.

Ahmad, K. and Henikoff, S. (2002). The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. Mol. Cell **9**: 1191–200.

Allison, L.A., Romaniuk, P.J., and Bakken, A.H. (1991). RNA-protein interactions of stored 5S RNA with TFIIIA and ribosomal protein L5 during Xenopus oogenesis. Dev. Biol. **144**: 129–144.

Alonso-Blanco, C. et al. (2016). 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. Cell **166**: 481–491.

Alves-Costa, F.A., Martins, C., Matos, F.D.C. de, Foresti, F., Oliveira, C., and Wasko, A.P. (2008). 5S rDNA characterization in twelve Sciaenidae fish species (Teleostei, Perciformes): depicting gene diversity and molecular markers. Genet. Mol. Biol. **31**: 303–307.

Bannister, A.J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. **21**: 381–95.

Barakat, A., Szick-Miranda, K., Chang, I.F., Guyot, R., Blanc, G., Cooke, R., Delseny, M., and Bailey-Serres, J. (2001). The organization of cytoplasmic ribosomal protein genes in the Arabidopsis genome. Plant Physiol. **127**: 398–415.

Barciszewska, M.Z., Szymanski, M., Erdmann, V.A., Barciszewski, J., Szymański, M., Erdmann, V.A., Barciszewski, J., Szymanski, M., Erdmann, V.A., and Barciszewski, J. (2000). 5S ribosomal RNA. Biomacromolecules **1**: 297–302.

Barski, A., Chepelev, I., Liko, D., Cuddapah, S., Fleming, A.B., Birch, J., Cui, K., White, R.J., and Zhao, K. (2010). Pol II and its associated epigenetic marks are present at pol III- transcribed non-coding RNA genes. Nat struct mol biol **17**: 629–634.

Bauwens, S., Van Oostveldt, P., Engler, G., and Van Montagu, M. (1991). Distribution of the rDNA and three classes of highly repetitive DNA in the chromatin of interphase nuclei of Arabidopsis thaliana. Chromosoma **101**: 41–48.

Benoit, M., Layat, E., Tourmente, S., and Probst, A. V (2013). Heterochromatin dynamics during developmental transitions in Arabidopsis - a focus on ribosomal DNA loci. Gene.

Bhargava, P. (2013). Epigenetic regulation of transcription by RNA polymerase III. Biochim. Biophys. Acta **1829**: 1015–25.

Bieker, J.J., Martin, P.L., and Roeder, R.G. (1985). Formation of a rate-limiting intermediate in 5S RNA gene transcription. Cell **40**: 119–27.

Birchler, J.A. and Veitia, R.A. (2012). Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proc. Natl. Acad. Sci. U. S. A. **109**: 14746–53.

Blevins, T., Pontes, O., Pikaard, C.S., and Meins, F. (2009). Heterochromatic siRNAs and DDM1 independently silence aberrant 5S rDNA transcripts in Arabidopsis. PLoS One **4**: e5932.

Bogenhagen, D.F. and Brown, D.D. (1981). Nucleotide sequences in Xenopus 5S DNA required for transcription termination. Cell **24**: 261–70.

Byrne, M.E. (2009). A role for the ribosome in development. Trends Plant Sci. **14**: 512–519.

Campell, B.R., Song, Y., Posch, T.E., Cullis, C.A., and Town, C.D. (1992). Sequence and organization of 5S ribosomal RNA-encoding genes of Arabidopsis thaliana. Gene **112**: 225–228.

Cartagena, J.A., Matsunaga, S., Seki, M., Kurihara, D., Yokoyama, M., Shinozaki, K., Fujimoto, S., Azumi, Y., Uchiyama, S., and Fukui, K. (2008). The Arabidopsis SDG4 contributes to the regulation of pollen tube growth by methylation of histone H3 lysines 4 and 36 in mature pollen. Dev. Biol. **315**: 355–368.

Cazaux, B., Catalan, J., Veyrunes, F., Douzery, E.J., and Britton-Davidian, J. (2011). Are ribosomal DNA clusters rearrangement hotspots?: a case study in the genus Mus (Rodentia, Muridae). BMC Evol. Biol. **11**: 124.

Chandler, J.W. (2008). Cotyledon organogenesis. J. Exp. Bot. **59**: 2917–31.

Chandrasekhara, C., Mohannath, G., Blevins, T., Pontvianne, F., and Pikaard, C.S. (2016). Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in Arabidopsis. Genes Dev. **30**: 177–90.

Chaubet, N., Clement, B., and Gigot, C. (1992). Genes encoding a histone H3.3-like variant in Arabidopsis contain intervening sequences. J. Mol. Biol. **225**: 569–574.

Chodavarapu, R.K. et al. (2010). Relationship between nucleosome positioning and DNA methylation. Nature **466**: 388–392.

Ciganda, M. and Williams, N. (2011). Eukaryotic 5S rRNA biogenesis. Wiley Interdiscip. Rev. RNA **2**: 523–33.

Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S., and Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. Nat. Nanotechnol. **4**: 265–70.

Cloix, C., Tutois, S., Mathieu, O., Cuvillier, C., Espagnol, M.C., Picard, G., and Tourmente, S. (2000). Analysis of 5S rDNA arrays in Arabidopsis thaliana: physical mapping and chromosome-specific polymorphisms. Genome Res. **10**: 679–690.

Cloix, C., Tutois, S., Yukawa, Y., Mathieu, O., Cuvillier, C., Espagnol, M.-C., Picard, G., and Tourmente, S. (2002). Analysis of the 5S RNA Pool in Arabidopsis thaliana: RNAs Are Heterogeneous and Only Two of the Genomic 5S Loci Produce Mature 5S RNA. Genome Res. **12**: 132–144.

Cloix, C., Yukawa, Y., Tutois, S., Sugiura, M., and Tourmente, S. (2003). In vitro analysis of the sequences required for transcription of the Arabidopsis thaliana 5S rRNA genes. Plant J. **35**: 251–261.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature **452**: 215–9.

Coleman-Derr, D. and Zilberman, D. (2012). DNA methylation, H2A.Z, and the regulation of constitutive expression. Cold Spring Harb. Symp. Quant. Biol. **77**: 147–54.

Condon, C., Squires, C., and Squires, C.L. (1995). Control of rRNA transcription in Escherichia coli. Microbiol. Rev. **59**: 623–45.

Copenhaver, G.P. and Pikaard, C.S. (1996). 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. Plant J. **9**: 273–82.

Creusot, F., Fouilloux, E., Dron, M., Lafleuriel, J., Picard, G., Billault, A., Le Paslier, D., Cohen, D., Chabouté, M.E., and Durr, A. (1995). The CIC library: a large insert YAC library for genome mapping in Arabidopsis thaliana. Plant J. cell Mol. Biol. **8**: 763–770.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol. **139**: 5–17.

Dammann, R., Lucchini, R., Koller, T., and Sogo, J.M. (1993). Chromatin structures and transcription of rDNA in yeast Saccharomyces cerevisiae. Nucleic Acids Res. **21**: 2331–8.

Davison, J., Tyagi, A., and Comai, L. (2007). Large-scale polymorphism of heterochromatic repeats in the DNA of Arabidopsis thaliana. BMC Plant Biol. **7**: 44.

Dieci, G., Fiorino, G., Castelnuovo, M., Teichmann, M., and Pagano, A. (2007). The expanding RNA polymerase III transcriptome. Trends Genet. **23**: 614–22.

Dileep, V., Rivera-Mulia, J.C., Sima, J., and Gilbert, D.M. (2015). Large-Scale Chromatin Structure-Function Relationships during the Cell Cycle and Development: Insights from Replication Timing. Cold Spring Harb. Symp. Quant. Biol. **80**: 53–63.

Dinman, J.D. (2005). 5S rRNA: Structure and Function from Head to Toe. Int. J. Biomed. Sci. **1**: 2–7.

Douet, J., Blanchard, B., Cuvillier, C., and Tourmente, S. (2008). Interplay of RNA Pol IV and ROS1 During Post-Embryonic 5S rDNA Chromatin Remodeling. Plant Cell Physiol. **49**: 1783–1791.

Douet, J., Tutois, S., and Tourmente, S. (2009a). A Pol V-mediated silencing, independent of RNA-directed DNA methylation, applies to 5S rDNA. PLoS Genet. **5**: e1000690.

Douet, J., Tutois, S., and Tourmente, S. (2009b). A Pol V–Mediated Silencing, Independent of RNA–Directed DNA Methylation, Applies to 5S rDNA. PLoS Genet. **5**: e1000690.

Du, J. et al. (2012). Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell **151**: 167–180.

Ebbs, M.L., Bartee, L., and Bender, J. (2005). H3 lysine 9 methylation is maintained on a transcribed inverted repeat by combined action of SUVH6 and SUVH4 methyltransferases. Mol. Cell. Biol. **25**: 10507–15.

Ebbs, M.L. and Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. Plant Cell **18**: 1166–76.

Eickbush, T.H. and Eickbush, D.G. (2007). Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics **175**: 477–85.

Eid, J. et al. (2009). Real-time DNA sequencing from single polymerase molecules. Science **323**: 133–8.

Engelke, D.R., Ng, S.Y., Shastry, B.S., and Roeder, R.G. (1980). Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell **19**: 717–728.

Fedoriw, A.M., Starmer, J., Yee, D., and Magnuson, T. (2012). Nucleolar association and transcriptional inhibition through 5S rDNA in mammals. PLoS Genet. **8**: e1002468.

Fransz, P. et al. (2016). Molecular, genetic and evolutionary analysis of a paracentric inversion in Arabidopsis thaliana. Plant J.

Fransz, P.F., Armstrong, S., Alonso-Blanco, C., Fischer, T.C., Torres-Ruiz, R.A., and Jones, G. (1998). Cytogenetics for the model system Arabidopsis thaliana. Plant J. **13**: 867–76.

Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T., and Jones, G.H. (2000). Integrated cytogenetic map of chromosome arm 4S of A. thaliana: structural organization of heterochromatic knob and centromere region. Cell **100**: 367–76.

Fransz, P.F. and de Jong, H. (2011). From nucleosome to chromosome: a dynamic organization of genetic information. Plant J. **66**: 4–17.

Fransz, P.F., De Jong, J.H., Lysak, M., Castiglione, M.R., and Schubert, I. (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. Proc. Natl. Acad. Sci. U. S. A. **99**: 14584–14589.

French, S.L., Osheim, Y.N., Cioci, F., Nomura, M., and Beyer, A.L. (2003). In exponentially growing Saccharomyces cerevisiae cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. Mol. Cell. Biol. **23**: 1558–68.

Fu, Y., Bannach, O., Chen, H., Teune, J.-H., Schmitz, A., Steger, G., Xiong, L., and Barbazuk, W.B. (2009). Alternative splicing of anciently exonized 5S rRNA regulates plant transcription factor TFIIIA. Genome Res. **19**: 913–21.

Garcia, S., Garnatje, T., Hidalgo, O., McArthur, E.D., Siljak-Yakovlev, S., and Vallès, J. (2007). Extensive ribosomal DNA (18S-5.8S-26S and 5S) colocalization in the North American endemic sagebrushes (subgenus Tridentatae, Artemisia, Asteraceae) revealed by FISH. Plant Syst. Evol. **267**: 79–92.

Garcia, S. and Kovařík, A. (2013). Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. Heredity (Edinb). **111**: 23–33.

Garcia, S., Lim, K.Y., Chester, M., Garnatje, T., Pellicer, J., Vallès, J., Leitch, A.R., and Kovarík, A. (2009). Linkage of 35S and 5S rRNA genes in Artemisia (family Asteraceae): first evidence from angiosperms. Chromosoma **118**: 85–97.

Garcia, S., Panero, J.L., Siroky, J., and Kovarik, A. (2010). Repeated reunions and splits feature the highly dynamic evolution of 5S and 35S ribosomal RNA genes (rDNA) in the Asteraceae family. BMC Plant Biol. **10**: 176.

Gibbons, J.G., Branco, A.T., Godinho, S.A., Yu, S., and Lemos, B. (2015). Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. Proc. Natl. Acad. Sci. U. S. A. **112**: 2485–90.

Goyon, C. (1996). Native DNA repeats and methylation in Ascobolus. Nucleic Acids Res. **24**: 3348–3356.

Grob, S., Schmid, M.W., and Grossniklaus, U. (2014). Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. Mol. Cell: 1–16.

Grob, S., Schmid, M.W., Luedtke, N.W., Wicker, T., and Grossniklaus, U. (2013). Characterization of chromosomal architecture in Arabidopsis by chromosome conformation capture. Genome Biol. **14**: R129.

Grummt, I. (2010). Wisely chosen paths--regulation of rRNA synthesis: delivered on 30 June 2010 at the 35th FEBS Congress in Gothenburg, Sweden. FEBS J. **277**: 4626–39.

Grummt, I. and Pikaard, C.S. (2003). Epigenetic silencing of RNA polymerase I transcription. Nat. Rev. Mol. Cell Biol. **4**: 641–9.

Guddat, U., Bakken, A.H., and Pieler, T. (1990). Protein-mediated nuclear export of RNA: 5S rRNA containing small RNPs in xenopus oocytes. Cell **60**: 619–28.

Guetg, C., Lienemann, P., Sirri, V., Grummt, I., Hernandez-Verdun, D., Hottiger, M.O., Fussenegger, M., and Santoro, R. (2010). The NoRC complex mediates heterochromatin formation and stability of silent rRNA genes and centromeric repeats. EMBO J. **29**: 2253–2253.

Guo, L., Yu, Y., Law, J.A., and Zhang, X. (2010). SET DOMAIN GROUP2 is the major histone H3 lysine [corrected] 4 trimethyltransferase in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **107**: 18557–62.

Hammond, M.C., Wachter, A., and Breaker, R.R. (2009). A plant 5S ribosomal RNA mimic regulates alternative splicing of transcription factor IIIA pre-mRNAs. Nat. Struct. Mol. Biol. **16**: 541–9.

Havlová, K., Dvořáčková, M., Peiro, R., Abia, D., Mozgová, I., Vansáčová, L., Gutierrez, C., and Fajkus, J. (2016). Variation of 45S rDNA intergenic spacers in Arabidopsis thaliana. Plant Mol. Biol.: 1–15.

He, X.-J., Chen, T., and Zhu, J.-K. (2011). Regulation and function of DNA methylation in plants and animals. Cell Res. **21**: 442–65.

Hennig, L., Taranto, P., Walser, M., Schönrock, N., and Gruissem, W. (2003). Arabidopsis MSI1 is required for epigenetic maintenance of reproductive development. Development **130**: 2555–65.

Henras, A.K., Plisson-Chastang, C., O'Donohue, M.-F., Chakraborty, A., and Gleizes, P.-E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. Wiley Interdiscip. Rev. RNA **6**: 225–42.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science **308**: 118–20.

Holmberg, L. and Nygård, O. (2000). Release of ribosome-bound 5S rRNA upon cleavage of the phosphodiester bond between nucleotides A54 and A55 in 5S rRNA. Biol. Chem. **381**: 1041–6.

Hummel, M., Cordewener, J.H.G., de Groot, J.C.M., Smeekens, S., America, A.H.P., and Hanson, J. (2012). Dynamic protein composition of Arabidopsis thaliana cytosolic ribosomes in response to sucrose feeding as revealed by label free MSE proteomics. Proteomics **12**: 1024–38.

Ide, S., Miyazaki, T., Maki, H., and Kobayashi, T. (2010). Abundance of ribosomal RNA gene copies maintains genome integrity. Science **327**: 693–6.

Ingouff, M., Hamamura, Y., Gourgues, M., Higashiyama, T., and Berger, F. (2007). Distinct Dynamics of HISTONE3 Variants between the Two Fertilization Products in Plants. Curr. Biol. **17**: 1032–1037.

Ingouff, M., Rademacher, S., Holec, S., Šoljić, L., Xin, N., Readshaw, A., Foo, S.H., Lahouze, B., Sprunck, S., and Berger, F. (2010). Zygotic Resetting of the HISTONE 3 Variant Repertoire Participates in Epigenetic Reprogramming in Arabidopsis. Curr. Biol. **20**: 2137–2143.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature **416**: 556–60.

Jacob, Y., Bergamin, E., Donoghue, M.T.A., Mongeon, V., LeBlanc, C., Voigt, P., Underwood, C.J., Brunzelle, J.S., Michaels, S.D., Reinberg, D., Couture, J.-F., and Martienssen, R.A. (2014). Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication. Science **343**: 1249–53.

Jacob, Y. and Feng, S. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat. Struct. … **16**: 763–768.

Jacob, Y., Stroud, H., Leblanc, C., Feng, S., Zhuo, L., Caro, E., Hassel, C., Gutierrez, C., Michaels, S.D., and Jacobsen, S.E. (2010). Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. Nature **466**: 987–991.

Jin, C. and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev. **21**: 1519–29.

John L. Woolford, J. and Warner, J.R. (1991). 10 The Ribosome and Its Synthesis. Cold Spring Harb. Monogr. Arch. Vol. 21A Vol. I Mol. Cell. Biol. Yeast Saccharomyces Genome Dyn. Protein Synth. Energ.

Karimi-Ashtiyani, R. et al. (2015). Point mutation impairs centromeric CENH3 loading and induces haploid plants. Proc. Natl. Acad. Sci.: 201504333.

Kassavetis, G.A., Bartholomew, B., Blanco, J.A., Johnson, T.E., and Geiduschek, E.P. (1991). Two essential components of the Saccharomyces cerevisiae transcription factor TFIIIB: transcription and DNA-binding properties. Proc. Natl. Acad. Sci. U. S. A. **88**: 7308–12.

Kaya, H., Shibahara, K.I., Taoka, K.I., Iwabuchi, M., Stillman, B., and Araki, T. (2001). FASCIATA genes for chromatin assembly factor-1 in arabidopsis maintain the cellular organization of apical meristems. Cell **104**: 131–42.

Khaitovich, P. and Mankin, A.S. (1999). Effect of antibiotics on large ribosomal subunit assembly reveals possible function of 5 S rRNA. J. Mol. Biol. **291**: 1025–34.

Kimura, H. and Cook, P.R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. J. Cell Biol. **153**: 1341–53.

Kobayashi, T. (2008). A new role of the rDNA and nucleolus in the nucleus--rDNA instability maintains genome integrity. Bioessays **30**: 267–72.

Kobayashi, T. (2011). Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. Cell. Mol. Life Sci. **68**: 1395–403.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell **128**: 693–705.

Kumar, S.V. and Wigge, P.A. (2010). H2A.Z-containing nucleosomes mediate the thermosensory response in Arabidopsis. Cell **140**: 136–47.

Kumekawa, N., Hosouchi, T., Tsuruoka, H., and Kotani, H. (2001). The size and sequence organization of the centromeric region of Arabidopsis thaliana chromosome 4. DNA Res. **8**: 285–290.

Kumekawa, N., Hosouchi, T., Tsuruoka, H., and Kotani, H. (2000). The size and sequence organization of the centromeric region of arabidopsis thaliana chromosome 5. DNA Res. **7**: 315–21.

Law, J.A. and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. **11**: 204–220.

Lawrence, R.J., Earley, K., Pontes, O., Silva, M., Chen, Z.J., Neves, N., Viegas, W., and Pikaard, C.S. (2004). A Concerted DNA Methylation/Histone Methylation Switch Regulates rRNA Gene Dosage Control and Nucleolar Dominance. Mol. Cell **13**: 599–609.

Layat, E., Cotterell, S., Vaillant, I., Yukawa, Y., Tutois, S., and Tourmente, S. (2012a). Transcript levels, alternative splicing and proteolytic cleavage of TFIIIA control 5S rRNA accumulation during Arabidopsis thaliana development. Plant J. **71**: 35–44.

Layat, E., Sáez-Vásquez, J., and Tourmente, S. (2012b). Regulation of Pol I-Transcribed 45S rDNA and Pol III-Transcribed 5S rDNA in Arabidopsis. Plant Cell Physiol. **53**: 267–276.

Lee, T.-J. et al. (2010). Arabidopsis thaliana chromosome 4 replicates in two phases that correlate with chromatin state. PLoS Genet. **6**: e1000982.

Lermontova, I., Koroleva, O., Rutten, T., Fuchs, J., Schubert, V., Moraes, I., Koszegi, D., and Schubert, I. (2011). Knockdown of CENH3 in Arabidopsis reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. Plant J. **68**: 40–50.

Leutwiler, L.S., Hough-Evans, B.R., and Meyerowitz, E.M. (1984). The DNA of Arabidopsis thaliana. MGG Mol. Gen. Genet. **194**: 15–23.

Long, Q. et al. (2013). Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nat. Genet. **45**: 884–890.

Loyola, A. and Almouzni, G. (2004). Histone chaperones, a supporting role in the limelight. Biochim. Biophys. Acta **1677**: 3–11.

Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000). DNA methylation in Drosophila melanogaster. Nature **408**: 538–40.

Mathieu, O., Jasencakova, Z., Vaillant, I., Gendrel, A.-V., Colot, V., Schubert, I., and Tourmente, S. (2003a). Changes in 5S rDNA Chromatin Organization and Transcription during Heterochromatin Establishment in Arabidopsis. Plant Cell **15**: 2929–2939.

Mathieu, O., Picard, G., and Tourmente, S. (2002a). Methylation of a euchromatin-heterochromatin transition region in Arabidopsis thaliana chromosome 5 left arm. Chromosom. Res. **10**: 455–466.

Mathieu, O., Probst, A. V, and Paszkowski, J. (2005). Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. EMBO J. **24**: 2783–91.

Mathieu, O., Yukawa, Y., Prieto, J.-L., Vaillant, I., Sugiura, M., and Tourmente, S. (2003b). Identification and characterization of transcription factor IIIA and ribosomal protein L5 from Arabidopsis thaliana. Nucleic Acids Res. **31**: 2424–33.

Mathieu, O., Yukawa, Y., Sugiura, M., Picard, G., and Tourmente, S. (2002b). 5S rRNA genes expression is not inhibited by DNA methylation in Arabidopsis. Plant J. **29**: 313–323.

McStay, B. and Grummt, I. (2008). The epigenetics of rRNA genes: from molecular to chromosome biology. Annu. Rev. Cell Dev. Biol. **24**: 131–57.

Mentewab, A.B., Jacobsen, M.J., and Flowers, R.A. (2011). Incomplete homogenization of 18 S ribosomal DNA coding regions in Arabidopsis thaliana. BMC Res. Notes **4**: 93.

Mozgová, I., Mokros, P., and Fajkus, J. (2010). Dysfunction of chromatin assembly factor 1 induces shortening of telomeres and loss of 45S rDNA in Arabidopsis thaliana. Plant Cell **22**: 2768–80.

Murata, M., Heslop-Harrison, J.S., and Motoyoshi, F. (1997). Physical mapping of the 5S ribosomal RNA genes in Arabidopsis thaliana by multi-color fluorescence in situ hybridization with cosmid clones. Plant J. cell Mol. Biol. **12**: 31–37.

Nagaki, K., Talbert, P.B., Zhong, C.X., Dawe, R.K., Henikoff, S., and Jiang, J. (2003). Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres. Genetics **163**: 1221–1225.

Naumann, U., Daxinger, L., Kanno, T., Eun, C., Long, Q., Lorkovic, Z.J., Matzke, M., and Matzke, A.J.M. (2011). Genetic evidence that DNA methyltransferase DRM2 has a direct catalytic role in RNA-directed DNA methylation in Arabidopsis thaliana. Genetics **187**: 977–9.

Ono, T., Kaya, H., Takeda, S., Abe, M., Ogawa, Y., Kato, M., Kakutani, T., Mittelsten Scheid, O., Araki, T., and Shibahara, K. (2006). Chromatin assembly factor 1 ensures the stable maintenance of silent chromatin states in Arabidopsis. Genes Cells **11**: 153–62.

Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell **120**: 613–22.

Ottoline Leyser, H.M. and Furner, I.J. (1992). Characterisation of three shoot apical meristem mutants of Arabidopsis thaliana. Development **116**: 397–403.

Paule, M.R. and White, R.J. (2000). Survey and summary: transcription by RNA polymerases I and III. Nucleic Acids Res. **28**: 1283–98.

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M.A., Fuchs, J., and Schubert, I. (2004). Chromosome territory arrangement and homologous pairing in nuclei of Arabidopsis thaliana are predominantly random except for NOR-bearing chromosomes. Chromosoma **113**: 258–269.

Pecinka, a., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N., and Scheid, O.M. (2010). Epigenetic Regulation of Repetitive Elements Is Attenuated by Prolonged Heat Stress in Arabidopsis. Plant Cell **22**: 3118–3129.

Perina, A., Seoane, D., González-Tizón, A.M., Rodríguez-Fariña, F., and Martínez-Lage, A. (2011). Molecular organization and phylogenetic analysis of 5S rDNA in crustaceans of the genus Pollicipes reveal birth-and-death evolution and strong purifying selection. BMC Evol. Biol. **11**: 304.

Phelps-Durr, T.L., Thomas, J., Vahab, P., and Timmermans, M.C.P. (2005). Maize rough sheath2 and its Arabidopsis orthologue ASYMMETRIC LEAVES1 interact with HIRA, a predicted histone chaperone, to maintain knox gene silencing and determinacy during organogenesis. Plant Cell **17**: 2886–98.

Pikaard, C.S. (2002). Transcription and tyranny in the nucleolus: the organization, activation, dominance and repression of ribosomal RNA genes. Arabidopsis Book **1**: e0083.

Pittman, R.H., Andrews, M.T., and Setzer, D.R. (1999). A feedback loop coupling 5 S rRNA synthesis to accumulation of a ribosomal protein. J. Biol. Chem. **274**: 33198–201.

Pontes, O., Costa-Nunes, P., Vithayathil, P., and Pikaard, C.S. (2009). RNA polymerase v functions in arabidopsis interphase heterochromatin organization independently of the 24-nt siRNA-directed DNA methylation pathway. Mol. Plant **2**: 700–710.

Pontes, O., Lawrence, R.J., Neves, N., Silva, M., Lee, J.-H., Chen, Z.J., Viegas, W., and Pikaard, C.S. (2003). Natural variation in nucleolar dominance reveals the relationship between nucleolus organizer chromatin topology and rRNA gene transcription in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. **100**: 11418–23.

Pontes, O., Li, C.F., Costa Nunes, P., Haag, J., Ream, T., Vitins, A., Jacobsen, S.E., and Pikaard, C.S. (2006). The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. Cell **126**: 79–92.

Pontvianne, F. et al. (2010). Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in Arabidopsis thaliana. PLoS Genet. **6**: e1001225.

Pontvianne, F., Blevins, T., Chandrasekhara, C., Feng, W., Stroud, H., Jacobsen, S.E., Michaels, S.D., and Pikaard, C.S. (2012). Histone methyltransferases regulating rRNA gene dose and dosage control in Arabidopsis. Genes Dev. **26**: 945–57.

Pontvianne, F., Blevins, T., Chandrasekhara, C., Mozgová, I., Hassel, C., Pontes, O.M.F., Tucker, S., Mokros, P., Muchová, V., Fajkus, J., and Pikaard, C.S. (2013). Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. Genes Dev. **27**: 1545–50.

Preuss, S. and Pikaard, C.S. (2007). rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. Biochim. Biophys. Acta **1769**: 383–92.

Probst, A. V, Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nat. Rev. Mol. Cell Biol. **10**: 192–206.

Probst, A. V, Fagard, M., Proux, F., Mourrain, P., Boutet, S., Earley, K., Lawrence, R.J., Pikaard, C.S., Murfett, J., Furner, I., Vaucheret, H., and Mittelsten Scheid, O. (2004). Arabidopsis histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. Plant Cell **16**: 1021–34.

Prokopowich, C.D., Gregory, T.R., and Crease, T.J. (2003). The correlation between rDNA copy number and genome size in eukaryotes. Genome / Natl. Res. Counc. Canada = Génome / Cons. Natl. Rech. Canada **46**: 48–50.

Ramirez-Parra, E. and Gutierrez, C. (2007). E2F regulates FASCIATA1, a chromatin assembly gene whose loss switches on the endocycle and activates gene expression by changing the epigenetic status. Plant Physiol. **144**: 105–120.

Raué, H.A. and Planta, R.J. (1991). Ribosome Biogenesis in Yeast. Prog. Nucleic Acid Res. Mol. Biol. **41**: 89–129.

Ray-Gallet, D., Woolfe, A., Vassias, I., Pellentz, C., Lacoste, N., Puri, A., Schultz, D.C., Pchelintsev, N.A., Adams, P.D., Jansen, L.E.T., and Almouzni, G. (2011). Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. Mol. Cell **44**: 928–41.

Rédei, G.P. (1992). A heuristic glance at the past of Arabidopsis genetics. Methods Arab. Res.: 1–15.

Richard, G.-F., Kerrest, A., and Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol. Mol. Biol. Rev. **72**: 686–727.

Rodriguez-Granados, N.Y., Ramirez-Prado, J.S., Veluchamy, A., Latrasse, D., Raynaud, C., Crespi, M., Ariel, F., and Benhamed, M. (2016). Put your 3D glasses on: plant chromatin is on show. J. Exp. Bot. **67**: 1–17.

Roudier, F. et al. (2011). Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J. **30**: 1928–38.

Schneeberger, R.G., Creissen, G.P., and Cullis, C.A. (1989). Chromosomal and molecular analysis of 5S RNA gene organization in the flax, Linum usitatissimum. Gene **83**: 75–84.

Schonrock, N., Exner, V., Probst, a., Gruissem, W., and Hennig, L. (2006). Functional Genomic Analysis of CAF-1 Mutants in Arabidopsis thaliana. J. Biol. Chem. **281**: 9560–9568.

Selker, E.U., Yanofsky, C., Driftmier, K., Metzenberg, R.L., Alzner-DeWeerd, B., and RajBhandary, U.L. (1981). Dispersed 5S RNA genes in N. crassa: structure, expression and evolution. Cell **24**: 819–28.

Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S.E., Bastolla, U., and Gutierrez, C. (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.

Setzer, D. and Brown, D. (1985). Formation and stability of the 5 S RNA transcription complex. J. Biol. Chem. **260**: 2483–2492.

Shaw, P.J. and Jordan, E.G. (1995). The nucleolus. Annu. Rev. Cell Dev. Biol. **11**: 93–121.

Shibata, F. and Murata, M. (2004). Differential localization of the centromere-specific proteins in the major centromeric satellite of Arabidopsis thaliana. J. Cell Sci. **117**: 2963–2970.

Shu, H., Nakamura, M., Siretskiy, A., Borghi, L., Moraes, I., Wildhaber, T., Gruissem, W., and Hennig, L. (2014). Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes. Genome Biol. **15**: R62.

Shu, H., Wildhaber, T., Siretskiy, A., Gruissem, W., and Hennig, L. (2012). Distinct modes of DNA accessibility in plant chromatin. Nat. Commun. **3**: 1281.

Simon, L., Voisin, M., Tatout, C., and Probst, A. V. (2015). Structure and Function of Centromeric and Pericentromeric Heterochromatin in Arabidopsis thaliana. Front. Plant Sci. **6**.

Smith, M.W., Meskauskas, A., Wang, P., Sergiev, P. V, and Dinman, J.D. (2001). Saturation mutagenesis of 5S rRNA in Saccharomyces cerevisiae. Mol. Cell. Biol. **21**: 8264–75.

Sone, T., Fujisawa, M., Takenaka, M., Nakagawa, S., Yamaoka, S., Sakaida, M., Nishiyama, R., Yamato, K.T., Ohmido, N., Fukui, K., Fukuzawa, H., and Ohyama, K. (1999). Bryophyte 5S rDNA was inserted into 45S rDNA repeat units after the divergence from higher land plants. Plant Mol. Biol. **41**: 679–85.

Steimer, A., Amedeo, P., Afsar, K., Fransz, P.F., Mittelsten Scheid, O., and Paszkowski, J. (2000). Endogenous targets of transcriptional gene silencing in Arabidopsis. Plant Cell **12**: 1165–78.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol. **21**: 64–72.

Stroud, H., Otero, S., Desvoyes, B., Ramirez-Parra, E., Jacobsen, S.E., and Gutierrez, C. (2012). Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. Proc. Natl. Acad. Sci. **109**: 5370–5375.

Stults, D.M., Killen, M.W., Williamson, E.P., Hourigan, J.S., Vargas, H.D., Arnold, S.M., Moscow, J.A., and Pierce, A.J. (2009). Human rRNA gene clusters are recombinational hotspots in cancer. Cancer Res. **69**: 9096–104.

Szick-Miranda, K. and Bailey-Serres, J. (2001). Regulated heterogeneity in 12-kDa P-protein phosphorylation and composition of ribosomes in maize (Zea mays L.). J. Biol. Chem. **276**: 10921–8.

Talbert, P.B. et al. (2012). A unified phylogeny-based nomenclature for histone variants. Epigenetics Chromatin **5**: 7.

Talbert, P.B. and Henikoff, S. (2010). Histone variants--ancient wrap artists of the epigenome. Nat. Rev. Mol. Cell Biol. **11**: 264–75.

Tamada, Y., Yun, J.-Y., Woo, S.C., and Amasino, R.M. (2009). ARABIDOPSIS TRITHORAX-RELATED7 is required for methylation of lysine 4 of histone H3 and for transcriptional activation of FLOWERING LOCUS C. Plant Cell **21**: 3257–69.

Tanurdzic, M., Vaughn, M.W., Jiang, H., Lee, T.-J., Slotkin, R.K., Sosinski, B., Thompson, W.F., Doerge, R.W., and Martienssen, R. a (2008). Epigenomic consequences of immortalized plant cell suspension culture. PLoS Biol. **6**: 2880–2895.

Tchurikov, N.A., Fedoseeva, D.M., Sosin, D. V, Snezhkina, A. V, Melnikova, N. V, Kudryavtseva, A. V, Kravatsky, Y. V, and Kretova, O. V (2015). Hot spots of DNA double-strand breaks and genomic contacts of human rDNA units are involved in epigenetic regulation. J. Mol. Cell Biol. **7**: 366–82.

Tessadori, F., Schulkes, R.K., Driel, R. Van, and Fransz, P. (2007). Light-regulated large-scale reorganization of chromatin during the floral transition in Arabidopsis. Plant J. **50**: 848–857.

Tessarz, P., Santos-Rosa, H., Robson, S.C., Sylvestersen, K.B., Nelson, C.J., Nielsen, M.L., and Kouzarides, T. (2014). Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. Nature **505**: 564–8.

Thompson, H.L., Schmidt, R., and Dean, C. (1996). Identification and distribution of seven classes of middle-repetitive DNA in the Arabidopsis thaliana genome. Nucleic Acids Res. **24**: 3017–22.

Torii, K.U., Mitsukawa, N., Oosumi, T., Matsuura, Y., Yokoyama, R., Whittier, R.F., and Komeda, Y. (1996). The Arabidopsis ERECTA gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. Plant Cell **8**: 735–746.

Tutois, S., Cloix, C., Cuvillier, C., Espagnol, M.C., Lafleuriel, J., Picard, G., and Tourmente, S. (1999). Structural analysis and physical mapping of a pericentromeric region of chromosome 5 of Arabidopsis thaliana. Chromosom. Res. **7**: 143–156.

Vahidi, H., Curran, J., Nelson, D.W., Webster, J.M., McClure, M.A., and Honda, B.M. (1988). Unusual sequences, homologous to 5S RNA, in ribosomal DNA repeats of the nematode Meloidogyne arenaria. J. Mol. Evol. **27**: 222–7.

Vaillant, I., Schubert, I., Tourmente, S., and Mathieu, O. (2006). MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. EMBO Rep. **7**: 1273–1278.

Vaillant, I., Tutois, S., Cuvillier, C., Schubert, I., and Tourmente, S. (2007). Regulation of Arabidopsis thaliana 5S rRNA Genes. Plant Cell Physiol. **48**: 745–752.

Vaillant, I., Tutois, S., Jasencakova, Z., Douet, J., Schubert, I., and Tourmente, S. (2008). Hypomethylation and hypermethylation of the tandem repetitive 5S rRNA genes in Arabidopsis. Plant J. cell Mol. Biol. **54**: 299–309.

Vaquero-Sedas, M.I. and Vega-Palas, M. a (2013). Differential association of Arabidopsis telomeres and centromeres with histone H3 variants. Sci. Rep. **3**: 1202.

Venkateswarlu, K., Lee, S.W., and Nazar, R.N. (1991). Conserved upstream sequence elements in plant 5S ribosomal RNA-encoding genes. Gene **105**: 249–54.

Waminal, N.E., Ryu, K.B., Park, B.R., and Kim, H.H. (2013). Phylogeny of Cucurbitaceae species in Korea based on 5S rDNA non-transcribed spacer. Genes Genomics **36**: 57–64.

Warmerdam, D.O., van den Berg, J., and Medema, R.H. (2016). Breaks in the 45S rDNA Lead to Recombination-Mediated Loss of Repeats. Cell Rep. **14**: 2519–27.

Weis, B.L., Kovacevic, J., Missbach, S., and Schleiff, E. (2015). Plant-Specific Features of Ribosome Biogenesis. Trends Plant Sci. **20**: 729–40.

White, R.J. (2011). Transcription by RNA Polymerase III: more compolex then we thought. Nat. Rev. Genet. **12**: 459–463.

Wicke, S., Costa, A., Muñoz, J., and Quandt, D. (2011). Restless 5S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. Mol. Phylogenet. Evol. **61**: 321–32.

Wollmann, H., Holec, S., Alden, K., Clarke, N.D., Jacques, P.-E., and Berger, F. (2012). Dynamic Deposition of Histone Variant H3.3 Accompanies Developmental Remodeling of the Arabidopsis Transcriptome. PLoS Genet. **8**.

Woo, H.R., Dittmer, T.A., and Richards, E.J. (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. PLoS Genet. **4**: e1000156.

Woo, H.R., Pontes, O., Pikaard, C.S., and Richards, E.J. (2007). VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. Genes Dev. **21**: 267–77.

Xu, L., Zhao, Z., Dong, A., Soubigou-Taconnat, L., Renou, J.-P., Steinmetz, A., and Shen, W.-H. (2008). Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in Arabidopsis thaliana. Mol. Cell. Biol. **28**: 1348–60.

Yelagandula, R. et al. (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in Arabidopsis. Cell **158**: 98–109.

Yelina, N.E., Lambing, C., Hardcastle, T.J., Zhao, X., Santos, B., and Henderson, I.R. (2015). DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. Genes Dev. **29**: 2183–202.

Yi, H., Sardesai, N., Fujinuma, T., Chan, C.-W., Veena, and Gelvin, S.B. (2006). Constitutive expression exposes functional redundancy between the Arabidopsis histone H2A gene HTA1 and other H2A gene family members. Plant Cell **18**: 1575–89.

Yuan, J., Adamski, R., and Chen, J. (2010). Focus on histone variant H2AX: to be or not to be. FEBS Lett. **584**: 3717–24.

van Zanten, M., Koini, M.A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., Koornneef, M., Fransz, P., and Soppe, W.J.J. (2011). Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. Proc. Natl. Acad. Sci. U. S. A. **108**: 20219–24.

Zapata, L., Ding, J., Willing, E., Hartwig, B., Bezdan, D., Jiao, W., and Patel, V. (2016). Chromosome-level assembly of Arabidopsis thaliana L er reveals the extent of translocation and inversion polymorphisms. **0**.

Zhang, T., Zhang, W., and Jiang, J. (2015). Genome-Wide Nucleosome Occupancy and Positioning and Their Impact on Gene Expression and Evolution in Plants. Plant Physiol. **168**: 1406–16.

Zhang, X., Bernatavichute, Y. V, Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2009). Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in Arabidopsis thaliana. Genome Biol. **10**: R62.

Zhao, Z., Yu, Y., Meyer, D., Wu, C., and Shen, W.-H. (2005). Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. Nat. Cell Biol. **7**: 1256–60.

Zhou, W., Zhu, Y., Dong, A., and Shen, W.H. (2015). Histone H2A/H2B chaperones: From molecules to chromatin-based functions in plant growth and development. Plant J.: 1–18.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature **456**: 125–9.

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

Annexes

I. **Revue centromères et péricentromères**

frontiers in Plant Science

MINI REVIEW
published: 30 November 2015
doi: 10.3389/fpls.2015.01049

Structure and Function of Centromeric and Pericentromeric Heterochromatin in *Arabidopsis thaliana*

*Lauriane Simon†, Maxime Voisin†, Christophe Tatout and Aline V. Probst**

CNRS UMR6293, INSERM U1103, Clermont University, GReD, Aubière, France

The centromere is a specific chromosomal region where the kinetochore assembles to ensure the faithful segregation of sister chromatids during mitosis and meiosis. Centromeres are defined by a local enrichment of the specific histone variant CenH3 mostly at repetitive satellite sequences. A larger pericentromeric region containing repetitive sequences and transposable elements surrounds the centromere that adopts a particular chromatin state characterized by specific histone variants and post-translational modifications and forms a transcriptionally repressive chromosomal environment. In the model organism *Arabidopsis thaliana* centromeric and pericentromeric domains form conspicuous heterochromatin clusters called chromocenters in interphase. Here we discuss, using *Arabidopsis* as example, recent insight into mechanisms involved in maintenance and establishment of centromeric and pericentromeric chromatin signatures as well as in chromocenter formation.

Keywords: centromere, chromocenter, histone variants, 3D nucleus, lamina, nuclear envelope

OPEN ACCESS

Edited by:
Inna Lermontova, Institute of Plant Genetics and Crop Plant Research, Germany

Reviewed by:
Célia Baroux, University of Zürich, Switzerland
Stefan Heckmann, University of Birmingham, UK

***Correspondence:**
Aline V. Probst
aline.probst@univ-bpclermont.fr

† *These authors have contributed equally to this work.*

Specialty section:
This article was submitted to Plant Cell Biology, a section of the journal Frontiers in Plant Science

Received: *29 August 2015*
Accepted: *09 November 2015*
Published: *30 November 2015*

Citation:
Simon L, Voisin M, Tatout C and Probst AV (2015) Structure and Function of Centromeric and Pericentromeric Heterochromatin in Arabidopsis thaliana. Front. Plant Sci. 6:1049. doi: 10.3389/fpls.2015.01049

Centromeres are essential chromosomal structures that were first defined as central restrictions of the mitotic chromosomes that function in chromosome segregation during cell division. Except for *Saccharomyces cerevisiae*, centromeres are not defined genetically by a specific DNA sequence but rather epigenetically by a particular chromatin environment and the presence of the specific histone variant CenH3. The centromeric and the surrounding pericentromeric chromosomal regions form heterochromatin domains that remain condensed during interphase (Heitz, 1928) and in some species like *Arabidopsis thaliana* these are clustered into chromocenter structures (**Figure 1**; Fransz et al., 2002). Here we discuss our current knowledge concerning sequence composition, chromatin features and interphase higher-order organization of centromeric and pericentromeric regions into chromocenters, referring to the centromeric region specifically as the part of the chromosome involved in kinetochore formation, while we refer to the pericentromeric domains as the adjacent chromatin regions (according to Gent and Dawe, 2011).

SEQUENCE COMPOSITION OF *ARABIDOPSIS* CENTROMERIC AND PERICENTROMERIC REGIONS

Centromeric sequences consist in most organisms of short repetitive DNA sequences arranged in tandem and/or transposable elements (Plohl et al., 2014). New centromeres can emerge from anonymous sequences but they gradually incorporate repetitive arrays (Han et al., 2009; Plohl et al., 2014) suggesting that DNA repeats may be a preferred DNA environment for centromere formation. In *A. thaliana*, centromeric regions mainly consist of a 178 bp long sequence also called AtCon, pAL1,

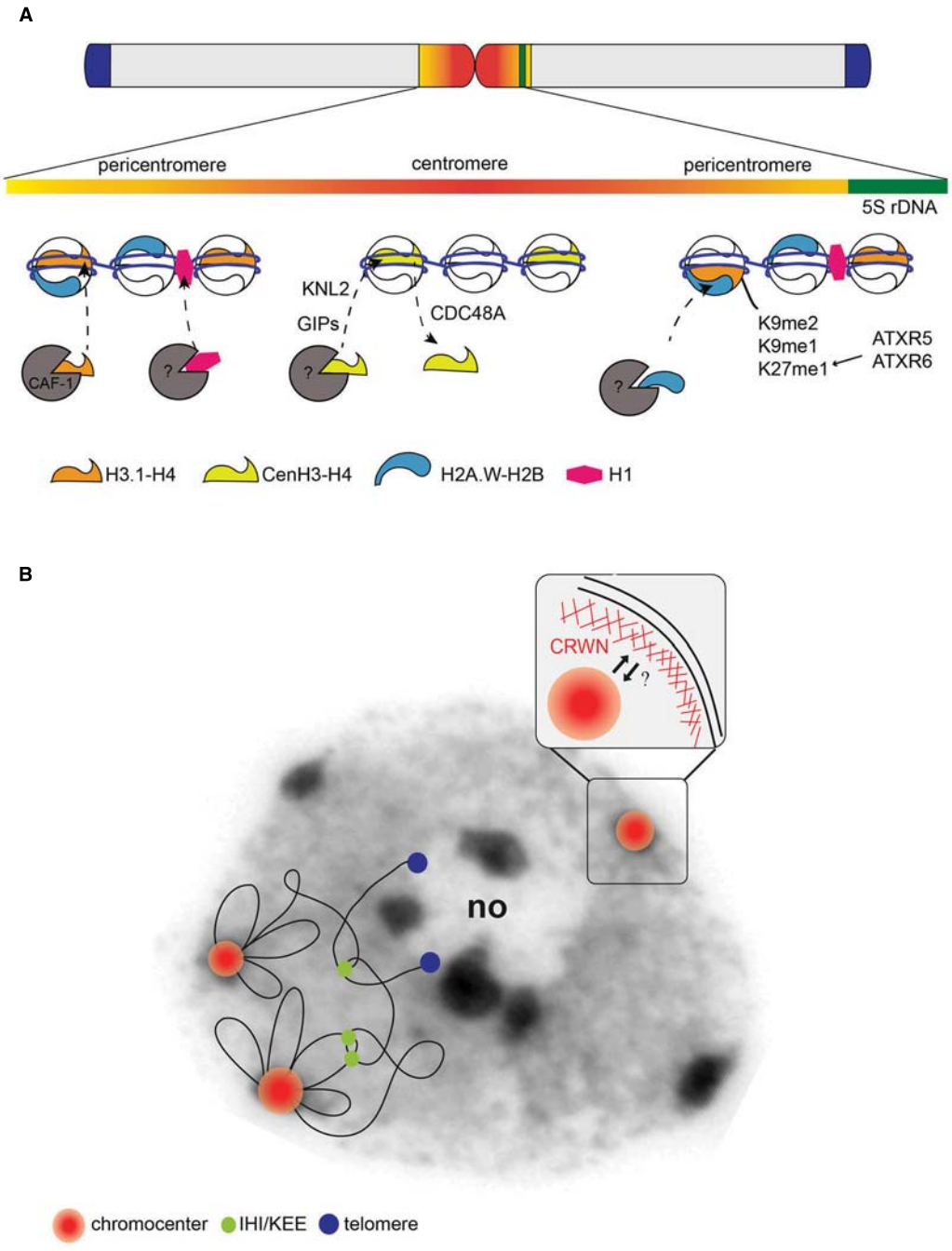


FIGURE 1 | (A) Model of an *Arabidopsis thaliana* chromosome. Enlargement of the central part of a chromosome (top) shows the sequence composition of the centromeric (red, mainly consisting of repetitive 180 bp repeats) and the pericentromeric region (orange, containing interspersed repeats, transposable elements and their derivatives) embedding a 5S rRNA gene locus (green) as well as their chromatin composition. The pericentromeric domains are enriched in the canonical histone H3.1 (orange), the H2A.W variant (blue) and the linker histone H1 (pink) as well as in repressive histone modifications such as H3K9me1/me2 and H3K27me1. The centromere is defined by enrichment in the H3 variant CenH3 (yellow). Whether CenH3 nucleosomes form large blocks or are interspersed with nucleosomes composed of canonical or specific variant histones and which histones are incorporated as placeholders during replication remains to be determined. CenH3, H1 and H2A.W are deposited by yet unknown histone chaperones. GIP1/GIP2 and KNL2 play a role in CenH3 deposition. In specific cell types, CenH3 is actively removed by the AAA-ATPase CDC48A. **(B)** Model of the organization of a chromosome in nuclear space overlaid on a DAPI stained image of an *Arabidopsis* leaf nucleus. Centromeric and pericentromeric regions of the 5 *Arabidopsis* chromosomes are tightly packed into chromocenters (red/orange). Chromocenters structure the chromosome in nuclear space by anchoring proximal euchromatic loops, while distal chromosomal regions tend to cluster with telomeres (blue) next to the nucleolus (no). Interactive heterochromatic island (IHI)/KNOT engaged element (KEE) regions, identified in Hi-C maps, form additional intra- and inter-chromosomal contacts (green). Enlargement at the top right shows the nuclear envelope coated internally by a lamina-like structure, which includes CRWN proteins. Chromocenters tend to localize at the nuclear periphery, but the physical link between nuclear envelope components or the nuclear lamina-like components remains to be identified.

or 180 bp (Kumekawa et al., 2000, 2001; Nagaki et al., 2003) that is arranged in head-to-tail tandem repeats ranging from ∼0.4 to 3 Mb (**Figure 1A**). The 180 bp repeats are interrupted by a 398 bp fragment of the Athila2 LTR called 106B (Thompson et al., 1996). While highly similar sequences are found on all chromosomes, some 180 bp repeat variants are specific to one centromere (Heslop-Harrison et al., 1999). Centromeres are flanked by a pericentromeric region, which in *Arabidopsis* contains various types of repeat sequences such as Athila retrotransposons, 500 bp and 160 bp repeats (Bauwens et al., 1991), "Transcriptional Silent Information" (TSI) composed of the 3′ half of an Athila retrotransposon (Steimer et al., 2000) and on some chromosomes the 5S rDNA clusters (Fransz et al., 1998; Cloix et al., 2000). Most repetitive sequences from the pericentromeric region and transposons are kept silent, while others like the 5S rRNA gene clusters (Cloix et al., 2002) are highly transcribed. Despite the efforts in genome assembly (Schneeberger et al., 2011), the establishment of the exact reference sequence for these highly repetitive chromosomal regions remains a challenge for the future.

THE CENTRAL ROLE OF CenH3 IN CENTROMERE DEFINITION

A high frequency of DNA repeats is not sufficient to define centromeres (Han et al., 2006; Birchler et al., 2010); instead, centromeres are determined by a specific chromatin environment. The basic subunit of chromatin, the nucleosome, comprises 146 bp of DNA that wrap around an octamer of core histones H3, H4, H2A, and H2B. A specific histone variant, called CenH3, replaces the canonical H3.1 in centromeric nucleosomes. CenH3 is enriched at 180 bp repeats as shown by Chromatin Immunoprecipitation (ChIP) and Fluorescence *in situ* hybridization (FISH) experiments (Nagaki et al., 2003; Shibata and Murata, 2004). In agreement with its central role in centromere definition, homozygous *cenH3* mutants are lethal and plants expressing RNAi constructs leading to reduced CenH3 levels show meiosis defects, partial sterility, and in older plants an increased 4C:2C ratio indicating G2 arrest (Lermontova et al., 2011). CenH3 proteins evolve rapidly, as example 23 out of 178 amino acids differ between the closely related species *A. thaliana* and *A. arenosa* (Talbert et al., 2002). The N-terminal tail is substantially longer compared to the canonical H3.1 or the variant H3.3 and particularly divergent between species, revealing adaptive evolution with the species-specific centromeric repeats (Talbert et al., 2002; Maheshwari et al., 2015). An additional domain involved in adaptive evolution is the histone fold domain including the loop 1 region that makes multiple contacts with DNA (Cooper and Henikoff, 2004). This histone fold domain has been found sufficient for CenH3 loading at centromeric sequences (Lermontova et al., 2006) and a single point mutation close to the loop1 region reduces CenH3 loading substantially (Karimi-Ashtiyani et al., 2015). While plants carrying this loading deficient CenH3 are fertile when selfed, backcrossing to wild type (WT) plants leads to haploid and aneuploid progeny, retaining only the WT CenH3. Unexpectedly, even CenH3 from distant monocotyledon species can complement *A. thaliana cenh3* mutants (Maheshwari et al., 2015). These results differ from other studies using N-or C-terminal GFP tagged versions of heterologous CenH3, where only tagged CenH3 from a closely related species was properly targeted or functionally complemented a *cenH3* mutant (Ravi et al., 2010). This shows that a GFP tagged CenH3 version is not functionally equivalent, as the large GFP tag may interfere with proper CenH3 loading or the assembly of kinetochore proteins.

Together, these studies underline a central role for CenH3 in centromere definition. Fast co-evolution of CenH3 and centromeric repeats is proposed to contribute to reproductive isolation and speciation (Ma et al., 2007; Plohl et al., 2014). Understanding these key mechanisms may have major application in breeding programs when interspecific crosses between cultivated species and their WT relatives are involved.

CenH3 DEPOSITION

CenH3 needs to be deposited in a controlled manner to avoid mis-localization to ectopic sites (Lacoste et al., 2014) that might seed neo-centromeres (Shang et al., 2013). Appropriate incorporation of CenH3 is therefore controlled by specific histone chaperones. In mammals, CenH3 is deposited post-mitotically (Jansen et al., 2007), while CenH3 nucleosome assembly takes place in G2 phase in *Arabidopsis* (Lermontova et al., 2006). In contrast to mitotic nuclei, meiosis includes also a post-divisional loading step during interkinesis (Schubert et al., 2014) and is associated with a specific loading pathway or quality check that eliminates modified CenH3 proteins (Lermontova et al., 2011; Ravi et al., 2011). No functional homolog of CenH3 chaperones known in humans (HJURP, Dunleavy et al., 2009; Foltz et al., 2009), Drosophila (CAL1, Chen et al., 2014) or yeast (SCM3, Camahort et al., 2007) has yet been identified in plants. In contrast, a homolog of yeast Mis18, which is implicated in forming the correct epigenetic context for CenH3 loading (Hayashi et al., 2004), has been identified in *Arabidopsis* and termed KINETOCHORE NULL2 (KNL2; Lermontova et al., 2013). KNL2 is mainly expressed in meristem tissues similar to CenH3 and except during mitosis localizes to centromeres during the whole cell cycle including G2 phase when CenH3 is loaded. Loss of KNL2 negatively impacts CenH3 expression and deposition (Lermontova et al., 2013), but also reduces DNA methylation and affects histone methyltransferase expression, suggesting that the chromatin context of centromeric or pericentromeric sequences repeats may play a role in CenH3 loading. Furthermore, recent work suggests a role for the γ-tubulin complex protein 3-interacting proteins (GIPs) in CenH3 loading or maintenance at centromeres (Batzenschlager et al., 2015). GIP proteins are found in a complex with CenH3 and a double *gip1 gip2* mutant shows reduced intensity of CenH3 signals, centromere cohesion defects and aneuploidy, despite increased levels of KNL2 in the *gip1 gip2* mutant background (Batzenschlager et al., 2015). Given that the CenH3 deposition machinery evolved rapidly and involves distinct players in yeast, Drosophila, mammals and plants, it can be speculated that GIP proteins are part of a plant-specific pathway contributing to CenH3 assembly.

Interestingly, CenH3 is associated only with specific subsets of 180 bp repeats (Shibata and Murata, 2004) and these are

hypomethylated (Zhang et al., 2008) compared to other subsets of 180 bp repeats that are hypermethylated, enriched in H3K9me2 and associated with the canonical histone H3.1 (Stroud et al., 2012; Wollmann et al., 2012; Vaquero-Sedas and Vega-Palas, 2013). To which extent CenH3 containing nucleosomes are interspersed with nucleosomes containing canonical H3.1 or its variant H3.3, and which of the CenH3 types is deposited as placeholder upon chromatin assembly during S-phase, remains to be elucidated in plants. Furthermore, the identification of the histone chaperone involved in CenH3 deposition and further characterization of the role of GIP proteins will be critical to better understand how CenH3 is specifically targeted to centromeric repeats.

While controlled CenH3 deposition is important, in some differentiated cells, CenH3 is also actively removed, such as in the vegetative pollen nucleus that does not contain visible CenH3 enrichment at centromeres compared to the sperm cell nuclei (Ingouff et al., 2009; Schoft et al., 2009). In the vegetative nucleus, CenH3 is sumoylated and removed by the AAA-ATPase molecular chaperone CDC48A to be targeted for proteolysis (Mérai et al., 2014).

THE PERICENTROMERIC REGION

The CenH3 containing centromere domain is flanked by pericentromeric heterochromatin that is highly DNA methylated, shows more regular nucleosome spacing than euchromatin (Chodavarapu et al., 2010) and is characterized by inaccessibility to DNAse I (Shu et al., 2012). Pericentromeric nucleosomes carry histone modifications repressive for transcription such as H4K20me1, H3K9me1, H3K9me2, and H3K27me1 (Tariq et al., 2003; Naumann et al., 2005; Fransz et al., 2006; Roudier et al., 2011; Shu et al., 2012; Sequeira-Mendes et al., 2014). As an example, H3K27me1 loss is associated with release of transcriptional silencing of TSI and certain transposons in the pericentromeric region (Jacob and Feng, 2009) as well as over-replication of pericentromeric sequences (Jacob et al., 2010). The histone-methyltransferases ATXR5 and ATXR6 preferentially mono-methylate lysine 27 of the canonical histone H3.1 (Jacob et al., 2014), which is highly enriched in pericentromeric regions (Vaquero-Sedas and Vega-Palas, 2013). Furthermore, plants deficient in the Chromatin Assembly Factor 1 (CAF-1) complex that deposits histone H3.1 in a replication-coupled manner in mammals (Smith and Stillman, 1989; Tagami et al., 2004) show stochastic reactivation of TSI and CACTA transposable elements (Takeda et al., 2004; Ono et al., 2006). Additional core histones also exist as specialized variants enriched in heterochromatin such as the H2A.W variants H2A.W.6, H2A.W.7 and H2A.W.12 (Yelagandula et al., 2014), which colocalize as RFP fusion proteins with H3K9me2 at pericentromeric regions by microscopy. Simultaneous loss of H3K9me2 enhances the phenotype of double h2a.w.6 h2a.w.7 mutants and leads to increased expression of certain transposons suggesting that histone/DNA methylation and H2A.W incorporation present two parallel pathways involved in heterochromatin maintenance. Therefore, an important role can be assigned to the incorporation of specific histone types in the establishment of the particular chromatin environment of the pericentromeric region. Canonical or histone variants affect chromatin organization both through their inherent physico-chemical properties and through their specific post-translational modifications that might be set in a nucleosomal context, e.g., H3K27me1 by ATXR5 and ATXR6 (Jacob et al., 2014), or already during assembly (Loyola et al., 2009) and synthesis (Rivera et al., 2015) of the respective histone as recently described in mammals. Despite the advances in the description of the pericentromeric heterochromatin signature, not much is known whether and how pericentromeric heterochromatin contributes to centromere function in plants. Chromatin or sequence features of the pericentromeric domain may play a role in loading of CenH3 at the centromere as it is the case in fission yeast (Folco et al., 2008; Catania et al., 2015) but this remains to be investigated.

ORGANIZATION OF CENTROMERIC AND PERICENTROMERIC CHROMATIN INTO CHROMOCENTERS

In interphase nuclei, FISH experiments indicated that centromeric and pericentromeric repeats cluster together in chromocenter structures (Fransz et al., 2002). Recent Hi-C analyses confirmed that repeated sequences are grouped together and revealed further intra and inter-chromosomal interactions, **Figure 1B**. Multiple reasons have been brought forward to explain the particular organization of centromeric and pericentromeric sequences into chromocenters: the clustering may compartmentalize silent chromatin away from euchromatin, help concentrate chromatin modifiers setting repressive chromatin marks or coordinate replication of this domain in time and space (Heitz, 1928; Quivy et al., 2004; Almouzni and Probst, 2011). Chromocenters are not randomly organized in nuclear space but instead preferentially localize into the most outer zone next to the nuclear periphery (Fransz et al., 2002; Fang and Spector, 2005; Andrey et al., 2010; Poulet et al., 2015). To date, there is no clear explanation for this preferential localization and several hypotheses can be proposed. First, this organization can be the result of non-specific forces acting on heterochromatin because of its elevated thickness and rigidity in respect to euchromatin (Cook and Marenduzzo, 2009; de Nooijer et al., 2009). Second, peripheral position may be advantageous to allow rapid contact between the centromere and microtubules at the beginning of cell division. Interestingly, GIP proteins have been shown to localize to both sides of the nuclear envelope and close to the chromocenters and may therefore be seen as good candidates to connect the microtubule machinery and the centromeres upon nuclear envelope breakdown (Batzenschlager et al., 2014). The identification of the structural components linking heterochromatin to the nuclear periphery is an active area of research in plants and the lamin-like structures including CRWN1-4 (CRoWded Nuclei) proteins are intriguing candidates (Dittmer et al., 2007; Fiserova et al., 2009; Goto et al., 2014). Indeed, *crwn1 crwn2* mutants show reduced nuclear volume and increased chromocenter clustering, while chromocenters are more dispersed in *crwn4* (Dittmer et al., 2007; Wang et al., 2013a; Poulet et al., 2015). Hi-C data in *crwn1* and *crwn4* mutants reveal higher chromosomal compaction

and increased interactions among pericentromeric regions, which reflects the altered chromocenter organization detected by FISH (Grob et al., 2014). Altered chromocenter organization was also observed in *syn4* and *cap-d3* mutants, lacking subunits of cohesin or condensin complexes respectively (Schubert et al., 2009, 2013). Given that both CRWN proteins and condensing/cohesion complexes affect chromocenter organization and chromosome compaction it might be interesting to further investigate whether a functional relationship exists between these complexes.

CHROMOCENTER MAINTENANCE AND DYNAMICS

Mutants impaired in factors involved in setting of epigenetic marks such as DNA methyltransferases (Soppe et al., 2002; Mathieu et al., 2007; Stroud et al., 2014) and histone K27 and K9 methyltransferases (Jacob and Feng, 2009; Yelagandula et al., 2014), in chromatin remodeling (Probst et al., 2003) or in chromatin assembly (Schönrock et al., 2006) affect heterochromatin organization in chromocenters. Furthermore, recent data suggest a role for the histone variant H2A.W in heterochromatin condensation into chromocenters based on its capacity to promote chromatin fiber-to-fiber interactions through its C-terminal end *in vitro* (Yelagandula et al., 2014) and accordingly *h2a.w* double or triple mutants show chromocenter decondensation. In addition to H2A.W, the linker histone H1 facilitates folding of the nucleosome into higher-order structures (Zhou et al., 2013 and references therein). The observation that some plant cells (such as the spore mother cells) show a drastic reduction in chromocenter compaction, concomitantly to H1 depletion (She et al., 2013; She and Baroux, 2015) suggests a role for the linker histone in pericentromeric chromatin organization, but a causal relationship remains to be established. During development, the organization of centromeric and pericentromeric sequences in chromocenters is dynamic (Benoit et al., 2013). For example during germination, chromocenter organization is lost 1 to 3 days after imbibition and only small and diffuse pre-chromocenters can be detected (van Zanten et al., 2011). Chromocenter assembly then takes place in cotyledons during a short time window between 3 and 5 days after germination (Mathieu et al., 2003; Douet et al., 2008; Bourbousse et al., 2015). Decondensation of chromocenters was also observed at later developmental stages such as during floral transition, when the plant undergoes reprogramming from vegetative to reproductive state, or during protoplast formation, which trigger a partial decondensation of 5S rDNA and 180 bp repeats (Tessadori et al., 2007a,b). Furthermore, the organization of chromocenters dynamically changes upon pathogen infection (Pavet et al., 2006), or under abiotic stresses (Probst and Mittelsten Scheid, 2015). As an example, chromocenters decondense during prolonged heat stress (Pecinka et al., 2010), which requires HEAT-INTOLERANT 4 (HIT4; Wang et al., 2013b). These dynamic changes in chromocenter organization during development or stress might reflect global chromatin changes, revealing the role of chromocenters in the organization of euchromatic loops in nuclear space thereby potentially contributing to gene expression regulation.

A ROLE FOR NON-CODING RNA IN CENTROMERE FUNCTION?

Centromeric and pericentromeric regions are essential for chromosome segregation in mitosis and meiosis and help to structure chromosomes through the formation of chromocenters in interphase. These are complex functions requesting many factors including specific DNA sequences, deposition of histone variants and epigenetic marks, as well as chromatin organization in nuclear space. While centromeric and pericentromeric regions form a generally repressive chromatin environment, some of these repetitive elements are expressed at low level in specific tissues or developmental stages and processed by the RNAi pathway (May et al., 2005; Slotkin et al., 2009; Slotkin, 2010). In recent years, evidence for RNA in centromere regulation and function accumulated in different organisms (reviewed in Gent et al., 2012). Examples include a role for non-coding RNAs in heterochromatin assembly in fission yeast (Volpe et al., 2002), HP1 recruitment (Maison et al., 2002) and chromocenter organization (Probst et al., 2010) in mammals, as well as CenH3 deposition in mammals and Drosophila (Quénet and Dalal, 2014; Rošić et al., 2014). Furthermore, the passage of RNA polymerase II itself is critical for centromere function (Catania et al., 2015; Chen et al., 2015). Understanding the complex interplay between DNA sequence, transcription, non-coding RNA, chromatin and nuclear environment in centromere function in plants will be a major challenge for the future.

AUTHOR CONTRIBUTIONS

LS and MV contributed equally to the content and to the drafting of the manuscript. CT and AP edited the manuscript. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

This work was supported by INSERM, CNRS and Clermont universities, by the French National Research Agency (ANR-11 JSV2 009 01, ANR-12-ISV6-0001 to AVP) and by the Region Auvergne and FEDER (Life GRID to CT). LS was supported by a PhD fellowship from the Region Auvergne.

REFERENCES

Almouzni, G., and Probst, A. V. (2011). Heterochromatin maintenance and establishment: lessons from the mouse pericentromere. *Nucleus* 2, 332–338. doi: 10.4161/nucl.2.5.17707

Andrey, P., Kiêu, K., Kress, C., Lehmann, G., Tirichine, L., Liu, Z., et al. (2010). Statistical analysis of 3D images detects regular spatial distributions of centromeres and chromocenters in animal and plant nuclei. *PLoS Comput. Biol.* 6:27. doi: 10.1371/journal.pcbi.1000853

Batzenschlager, M., Herzog, E., Houlné, G., Schmit, A.-C., and Chabouté, M.-E. (2014). GIP/MZT1 proteins orchestrate nuclear shaping. *Front. Plant Sci.* 5:29. doi: 10.3389/fpls.2014.00029

Batzenschlager, M., Lermontova, I., Schubert, V., Fuchs, J., Berr, A., Koini, M. A., et al. (2015). *Arabidopsis* MZT1 homologs GIP1 and GIP2 are essential

for centromere architecture. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8656–8660. doi: 10.1073/pnas.1506351112

Bauwens, S., Van Oostveldt, P., Engler, G., and Van Montagu, M. (1991). Distribution of the rDNA and three classes of highly repetitive DNA in the chromatin of interphase nuclei of *Arabidopsis thaliana*. *Chromosoma* 101, 41–48.

Benoit, M., Layat, E., Tourmente, S., and Probst, A. V. (2013). Heterochromatin dynamics during developmental transitions in *Arabidopsis*—a focus on ribosomal DNA loci. *Gene* 526, 39–45. doi: 10.1016/j.gene.2013.01.060

Birchler, J. A., Krishnaswamy, L., Gaeta, R. T., Masonbrink, R. E., and Zhao, C. (2010). Engineered minichromosomes in plants. *CRC. Crit. Rev. Plant Sci.* 29, 135–147. doi: 10.1080/07352681003709918

Bourbousse, C., Mestiri, I., Zabulon, G., Bourge, M., Formiggini, F., Koini, M. A., et al. (2015). Light signaling controls nuclear architecture reorganization during seedling establishment. *Proc. Natl. Acad. Sci. U.S.A.* 112, E2836–E2844. doi: 10.1073/pnas.1503512112

Camahort, R., Li, B., Florens, L., Swanson, S. K., Washburn, M. P., and Gerton, J. L. (2007). Scm3 is essential to recruit the histone h3 variant cse4 to centromeres and to maintain a functional kinetochore. *Mol. Cell* 26, 853–865. doi: 10.1016/j.molcel.2007.05.013

Catania, S., Pidoux, A. L., and Allshire, R. C. (2015). Sequence features and transcriptional stalling within centromere DNA promote establishment of CENP-A chromatin. *PLoS Genet.* 11:e1004986. doi: 10.1371/journal.pgen.1004986

Chen, C.-C., Bowers, S., Lipinszki, Z., Palladino, J., Trusiak, S., Bettini, E., et al. (2015). Establishment of centromeric chromatin by the CENP-A assembly factor CAL1 requires FACT-mediated transcription. *Dev. Cell* 34, 73–84. doi: 10.1016/j.devcel.2015.05.012

Chen, C. C., Dechassa, M. L., Bettini, E., Ledoux, M. B., Belisario, C., Heun, P., et al. (2014). CAL1 is the Drosophila CENP-A assembly factor. *J. Cell Biol.* 204, 313–329. doi: 10.1083/jcb.201305036

Chodavarapu, R. K., Feng, S., Bernatavichute, Y. V., Chen, P.-Y., Stroud, H., Yu, Y., et al. (2010). Relationship between nucleosome positioning and DNA methylation. *Nature* 466, 388–392. doi: 10.1038/nature09147

Cloix, C., Tutois, S., Mathieu, O., Cuvillier, C., Espagnol, M. C., Picard, G., et al. (2000). Analysis of 5S rDNA arrays in *Arabidopsis thaliana*: physical mapping and chromosome-specific polymorphisms. *Genome Res.* 10, 679–690. doi: 10.1101/gr.10.5.679

Cloix, C., Tutois, S., Yukawa, Y., Mathieu, O., Cuvillier, C., Espagnol, M.-C., et al. (2002). Analysis of the 5S RNA pool in *Arabidopsis thaliana*: RNAs Are heterogeneous and only two of the genomic 5S loci produce mature 5S RNA. *Genome Res.* 12, 132–144. doi: 10.1101/gr.181301

Cook, P. R., and Marenduzzo, D. (2009). Entropic organization of interphase chromosomes. *J. Cell Biol.* 186, 825–834. doi: 10.1083/jcb.200903083

Cooper, J. L., and Henikoff, S. (2004). Adaptive evolution of the histone fold domain in centromeric histones. *Mol. Biol. Evol.* 21, 1712–1718. doi: 10.1093/molbev/msh179

de Nooijer, S., Wellink, J., Mulder, B., and Bisseling, T. (2009). Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei. *Nucleic Acids Res.* 37, 3558–3568. doi: 10.1093/nar/gkp219

Dittmer, T. A., Stacey, N. J., Sugimoto-Shirasu, K., and Richards, E. J. (2007). LITTLE NUCLEI genes affecting nuclear morphology in *Arabidopsis thaliana*. *Plant Cell* 19, 2793–2803. doi: 10.1105/tpc.107.053231

Douet, J., Blanchard, B., Cuvillier, C., and Tourmente, S. (2008). Interplay of RNA Pol IV and ROS1 during post-embryonic 5S rDNA chromatin remodeling. *Plant Cell Physiol.* 49, 1783–1791. doi: 10.1093/pcp/pcn152

Dunleavy, E. M., Roche, D., Tagami, H., Lacoste, N., Ray-Gallet, D., Nakamura, Y., et al. (2009). HJURP Is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres. *Cell* 137, 485–497. doi: 10.1016/j.cell.2009.02.040

Fang, Y., and Spector, D. L. (2005). Centromere positioning and dynamics in living *Arabidopsis* plants. *Mol. Biol. Cell* 16, 5710–5718. doi: 10.1091/mbc.E05-08-0706

Fiserova, J., Kiseleva, E., and Goldberg, M. W. (2009). Nuclear envelope and nuclear pore complex structure and organization in tobacco BY-2 cells. *Plant J.* 59, 243–255. doi: 10.1111/j.1365-313X.2009.03865.x

Folco, H. D., Pidoux, A. L., Urano, T., and Allshire, R. C. (2008). Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. *Science* 319, 94–97. doi: 10.1126/science.1150944

Foltz, D. R., Jansen, L. E. T., Bailey, A. O., Yates, J. R., Bassett, E. A., Wood, S., et al. (2009). Centromere-specific assembly of CENP-A nucleosomes is mediated by HJURP. *Cell* 137, 472–484. doi: 10.1016/j.cell.2009.02.039

Fransz, P. F., Armstrong, S., Alonso-Blanco, C., Fischer, T. C., Torres-Ruiz, R. A., and Jones, G. (1998). Cytogenetics for the model system *Arabidopsis thaliana*. *Plant J.* 13, 867–876.

Fransz, P. F., De Jong, J. H., Lysak, M., Castiglione, M. R., and Schubert, I. (2002). Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14584–14589. doi: 10.1073/pnas.212325299

Fransz, P. F., ten Hoopen, R., and Tessadori, F. (2006). Composition and formation of heterochromatin in *Arabidopsis thaliana*. *Chromosom. Res.* 14, 71–82. doi: 10.1007/s10577-005-1022-5

Gent, J. I., and Dawe, R. K. (2011). RNA as a structural and regulatory component of the centromere. *Annu. Rev. Genet.* 46, 443–453. doi: 10.1146/annurev-genet-110711-155419

Gent, J. I., Dong, Y., Jiang, J., and Dawe, R. K. (2012). Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications. *Nucleic Acids Res.* 40, 1550–1560. doi: 10.1093/nar/gkr862

Goto, C., Tamura, K., Fukao, Y., Shimada, T., and Hara-Nishimura, I. (2014). The novel nuclear envelope protein KAKU4 modulates nuclear morphology in *Arabidopsis*. *Plant Cell* 26, 2143–2155. doi: 10.1105/tpc.113.122168

Grob, S., Schmid, M. W., and Grossniklaus, U. (2014). Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of drosophila. *Mol. Cell* 55, 678–679. doi: 10.1016/j.molcel.2014.07.009

Han, F., Lamb, J. C., and Birchler, J. A. (2006). High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3238–3243. doi: 10.1073/pnas.0509650103

Han, Y., Zhang, Z., Liu, C., Liu, J., Huang, S., Jiang, J., et al. (2009). Centromere repositioning in cucurbit species: implication of the genomic impact from centromere activation and inactivation. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14937–14941. doi: 10.1073/pnas.0904833106

Hayashi, T., Fujita, Y., Iwasaki, O., Adachi, Y., Takahashi, K., and Yanagida, M. (2004). Mis16 and Mis18 are required for CENP-A loading and histone deacetylation at centromeres. *Cell* 118, 715–729. doi: 10.1016/j.cell.2004.09.002

Heitz, E. (1928). Das heterochromatin der Moose. *Jahrb Wiss Bot.* 69, 762–818.

Heslop-Harrison, J. S., Murata, M., Ogura, Y., Schwarzacher, T., and Motoyoshi, F. (1999). Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes. *Plant Cell* 11, 31–42. doi: 10.1105/tpc.11.1.31

Ingouff, M., Sakata, T., Li, J., Sprunck, S., Dresselhaus, T., and Berger, F. (2009). The two male gametes share equal ability to fertilize the egg cell in *Arabidopsis thaliana*. *Curr. Biol.* 19, 19–20. doi: 10.1016/j.cub.2008.11.025

Jacob, Y., Bergamin, E., Donoghue, M. T. A., Mongeon, V., LeBlanc, C., Voigt, P., et al. (2014). Selective methylation of histone H3 variant H3.1 regulates heterochromatin replication. *Science* 343, 1249–1253. doi: 10.1126/science.1248357

Jacob, Y., and Feng, S. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat. Struct.* 16, 763–768. doi: 10.1038/nsmb.1611

Jacob, Y., Stroud, H., Leblanc, C., Feng, S., Zhuo, L., Caro, E., et al. (2010). Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* 466, 987–991. doi: 10.1038/nature09290

Jansen, L. E. T., Black, B. E., Foltz, D. R., and Cleveland, D. W. (2007). Propagation of centromeric chromatin requires exit from mitosis. *J. Cell Biol.* 176, 795–805. doi: 10.1083/jcb.200701066

Karimi-Ashtiyani, R., Ishii, T., Niessen, M., Stein, N., Heckmann, S., Gurushidze, M., et al. (2015). Point mutation impairs centromeric CENH3 loading and induces haploid plants. *Proc. Natl. Acad. Sci. U.S.A.* 201504333. doi: 10.1073/pnas.1504333112

Kumekawa, N., Hosouchi, T., Tsuruoka, H., and Kotani, H. (2000). The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 5. *DNA Res.* 7, 315–321. doi: 10.1093/dnares/7.6.315

Kumekawa, N., Hosouchi, T., Tsuruoka, H., and Kotani, H. (2001). The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 4. *DNA Res.* 8, 285–290. doi: 10.1093/dnares/8.6.285

Lacoste, N., Woolfe, A., Tachiwana, H., Garea, A. V., Barth, T., Cantaloube, S., et al. (2014). Mislocalization of the centromeric histone variant CenH3/CENP-A

in human cells depends on the chaperone DAXX. *Mol. Cell* 53, 631–644. doi: 10.1016/j.molcel.2014.01.018

Lermontova, I., Koroleva, O., Rutten, T., Fuchs, J., Schubert, V., Moraes, I., et al. (2011). Knockdown of CENH3 in *Arabidopsis* reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. *Plant J.* 68, 40–50. doi: 10.1111/j.1365-313X.2011.04664.x

Lermontova, I., Kuhlmann, M., Friedel, S., Rutten, T., Heckmann, S., Sandmann, M., et al. (2013). *Arabidopsis* kinetochore null2 is an upstream component for centromeric histone H3 variant cenH3 deposition at centromeres. *Plant Cell* 25, 3389–3404. doi: 10.1105/tpc.113.114736

Lermontova, I., Schubert, V., Fuchs, J., Klatte, S., Macas, J., and Schubert, I. (2006). Loading of *Arabidopsis* centromeric histone CENH3 occurs mainly during G2 and requires the presence of the histone fold domain. *Plant Cell* 18, 2443–2451. doi: 10.1105/tpc.106.043174

Loyola, A., Tagami, H., Bonaldi, T., Roche, D., Quivy, J. P., Imhof, A., et al. (2009). The HP1alpha-CAF1-SetDB1-containing complex provides H3K9me1 for Suv39-mediated K9me3 in pericentric heterochromatin. *EMBO Rep.* 10, 769–775. doi: 10.1038/embor.2009.90

Ma, J., Wing, R. A., Bennetzen, J. L., and Jackson, S. A. (2007). Plant centromere organization: a dynamic structure with conserved functions. *Trends Genet.* 23, 134–139. doi: 10.1016/j.tig.2007.01.004

Maheshwari, S., Tan, E. H., West, A., Franklin, F. C. H., Comai, L., and Chan, S. W. L. (2015). Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. *PLoS Genet.* 11:e1004970. doi: 10.1371/journal.pgen.1004970

Maison, C., Bailly, D., Peters, A. H. F. M., Quivy, J.-P., Roche, D., Taddei, A., et al. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.* 30, 329–334. doi: 10.1038/ng843

Mathieu, O., Jasencakova, Z., Vaillant, I., Gendrel, A.-V., Colot, V., Schubert, I., et al. (2003). Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in *Arabidopsis*. *Plant Cell* 15, 2929–2939. doi: 10.1105/tpc.017467

Mathieu, O., Reinders, J., Čaikovski, M., Smathajitt, C., and Paszkowski, J. (2007). Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG methylation. *Cell* 130, 851–862. doi: 10.1016/j.cell.2007.07.007

May, B. P., Lippman, Z. B., Fang, Y., Spector, D. L., and Martienssen, R. a. (2005). Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats. *PLoS Genet.* 1:e79. doi: 10.1371/journal.pgen.0010079

Mérai, Z., Chumak, N., García-Aguilar, M., Hsieh, T.-F., Nishimura, T., Schoft, V. K., et al. (2014). The AAA-ATPase molecular chaperone Cdc48/p97 disassembles sumoylated centromeres, decondenses heterochromatin, and activates ribosomal RNA genes. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16166–16171. doi: 10.1073/pnas.1418564111

Nagaki, K., Talbert, P. B., Zhong, C. X., Dawe, R. K., Henikoff, S., and Jiang, J. (2003). Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. *Genetics* 163, 1221–1225.

Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., et al. (2005). Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in *Arabidopsis*. *EMBO J.* 24, 1418–1429. doi: 10.1038/sj.emboj.7600604

Ono, T., Kaya, H., Takeda, S., Abe, M., Ogawa, Y., Kato, M., et al. (2006). Chromatin assembly factor 1 ensures the stable maintenance of silent chromatin states in *Arabidopsis*. *Genes Cells* 11, 153–162. doi: 10.1111/j.1365-2443.2006.00928.x

Pavet, V., Quintero, C., Cecchini, N. M., Rosa, A. L., and Alvarez, M. E. (2006). *Arabidopsis* displays centromeric DNA hypomethylation and cytological alterations of heterochromatin upon attack by *Pseudomonas syringae*. *Mol. Plant Microbe Interact.* 19, 577–587. doi: 10.1094/MPMI-19-0577

Pecinka, A., Dinh, H. Q., Baubec, T., Rosa, M., Lettner, N., and Scheid, O. M. (2010). Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in *Arabidopsis*. *Plant Cell* 22, 3118–3129. doi: 10.1105/tpc.110.078493

Plohl, M., Meštrović, N., and Mravinac, B. (2014). Centromere identity from the DNA point of view. *Chromosoma* 123, 313–325. doi: 10.1007/s00412-014-0462-0

Poulet, A., Arganda-Carreras, I., Legland, D., Probst, A. V., Andrey, P., and Tatout, C. (2015). NucleusJ: an ImageJ plugin for quantifying 3D images of interphase nuclei. *Bioinformatics* 31, 1144–1146. doi: 10.1093/bioinformatics/btu774

Probst, A. V., Fransz, P. F., Paszkowski, J., and Mittelsten Scheid, O. (2003). Two means of transcriptional reactivation within heterochromatin. *Plant J. cell Mol. Biol.* 33, 743–749. doi: 10.1046/j.1365-313X.2003.01667.x

Probst, A. V., and Mittelsten Scheid, O. (2015). Stress-induced structural changes in plant chromatin. *Curr. Opin. Plant Biol.* 27, 8–16. doi: 10.1016/j.pbi.2015.05.011

Probst, A. V., Okamoto, I., Casanova, M., El Marjou, F., Le Baccon, P., and Almouzni, G. (2010). A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. *Dev. Cell* 19, 625–638. doi: 10.1016/j.devcel.2010.09.002

Quénet, D., and Dalal, Y. (2014). A long non-coding RNA is required for targeting centromeric protein A to the human centromere. *Elife* 3, e03254. doi: 10.7554/eLife.03254

Quivy, J.-P., Roche, D., Kirschner, D., Tagami, H., Nakatani, Y., and Almouzni, G. (2004). A CAF-1 dependent pool of HP1 during heterochromatin duplication. *EMBO J.* 23, 3516–3526. doi: 10.1038/sj.emboj.7600362

Ravi, M., Kwong, P. N., Menorca, R. M. G., Valencia, J. T., Ramahi, J. S., Stewart, J. L., et al. (2010). The rapidly evolving centromere-specific histone has stringent functional requirements in *Arabidopsis thaliana*. *Genetics* 186, 461–471. doi: 10.1534/genetics.110.120337

Ravi, M., Shibata, F., Ramahi, J. S., Nagaki, K., Chen, C., Murata, M., et al. (2011). Meiosis-specific loading of the centromere-specific histone CENH3 in *Arabidopsis thaliana*. *PLoS Genet.* 7:e1002121. doi: 10.1371/journal.pgen.1002121

Rivera, C., Saavedra, F., Alvarez, F., Díaz-Celis, C., Ugalde, V., Li, J., et al. (2015). Methylation of histone H3 lysine 9 occurs during translation. *Nucleic Acids Res.* 43, 9097–9106. doi: 10.1093/nar/gkv929

Rošić, S., Köhler, F., and Erhardt, S. (2014). Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. *J. Cell Biol.* 207, 335–349. doi: 10.1083/jcb.201404097

Roudier, F., Ahmed, I., Bérard, C., Sarazin, A., Mary-Huard, T., Cortijo, S., et al. (2011). Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J.* 30, 1928–1938. doi: 10.1038/emboj.2011.103

Schneeberger, K., Ossowski, S., Ott, F., Klein, J. D., Wang, X., Lanz, C., et al. (2011). Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10249–10254. doi: 10.1073/pnas.1107739108

Schoft, V. K., Chumak, N., Mosiolek, M., Slusarz, L., Komnenovic, V., Brownfield, L., et al. (2009). Induction of RNA-directed DNA methylation upon decondensation of constitutive heterochromatin. *EMBO Rep.* 10, 1015–1021. doi: 10.1038/embor.2009.152

Schönrock, N., Exner, V., Probst, A., Gruissem, W., and Hennig, L. (2006). Functional genomic analysis of CAF-1 mutants in *Arabidopsis thaliana*. *J. Biol. Chem.* 281, 9560–9568. doi: 10.1074/jbc.M513426200

Schubert, V., Lermontova, I., and Schubert, I. (2013). The *Arabidopsis* CAP-D proteins are required for correct chromatin organisation, growth and fertility. *Chromosoma* 122, 517–533. doi: 10.1007/s00412-013-0424-y

Schubert, V., Lermontova, I., and Schubert, I. (2014). Loading of the centromeric histone H3 variant during meiosis-how does it differ from mitosis? *Chromosoma* 123, 491–497. doi: 10.1007/s00412-014-0466-9

Schubert, V., Weißleder, A., Ali, H., Fuchs, J., Lermontova, I., Meister, A., et al. (2009). Cohesin gene defects may impair sister chromatid alignment and genome stability in *Arabidopsis thaliana*. *Chromosoma* 118, 591–605. doi: 10.1007/s00412-009-0220-x

Sequeira-Mendes, J., Aragüez, I., Peiró, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S. E., 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

Shang, W.-H., Hori, T., Martins, N. M. C., Toyoda, A., Misu, S., Monma, N., et al. (2013). Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. *Dev. Cell* 24, 635–648. doi: 10.1016/j.devcel.2013.02.009

She, W., and Baroux, C. (2015). Chromatin dynamics in pollen mother cells underpin a common scenario at the somatic-to-reproductive fate transition of both the male and female lineages in *Arabidopsis*. *Front. Plant Sci.* 6:294. doi: 10.3389/fpls.2015.00294

She, W., Grimanelli, D., Rutowicz, K., Whitehead, M. W. J., Puzio, M., Kotlinski, M., et al. (2013). Chromatin reprogramming during the somatic-to-reproductive cell fate transition in plants. *Development* 140, 4008–4019. doi: 10.1242/dev.095034

Shibata, F., and Murata, M. (2004). Differential localization of the centromere-specific proteins in the major centromeric satellite of *Arabidopsis thaliana*. *J. Cell Sci.* 117, 2963–2970. doi: 10.1242/jcs.01144

Shu, H., Wildhaber, T., Siretskiy, A., Gruissem, W., and Hennig, L. (2012). Distinct modes of DNA accessibility in plant chromatin. *Nat. Commun.* 3, 1281. doi: 10.1038/ncomms2259

Slotkin, R. K. (2010). The epigenetic control of the athila family of retrotransposons in *Arabidopsis*. *Epigenetics* 5, 483–490. doi: 10.4161/epi.5.6.12119

Slotkin, R. K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J. D., Feijó, J. A., et al. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* 136, 461–472. doi: 10.1016/j.cell.2008.12.038

Smith, S., and Stillman, B. (1989). Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication *in vitro*. *Cell* 58, 15–25. doi: 10.1016/0092-8674(89)90398-X

Soppe, W. J. J., Jasencakova, Z., Houben, A., Kakutani, T., Meister, A., Huang, M. S., et al. (2002). DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO J.* 21, 6549–6559. doi: 10.1093/emboj/cdf657

Steimer, A., Amedeo, P., Afsar, K., Fransz, P. F., Mittelsten Scheid, O., and Paszkowski, J. (2000). Endogenous targets of transcriptional gene silencing in *Arabidopsis*. *Plant Cell* 12, 1165–1178. doi: 10.1105/tpc.12.7.1165

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., et al. (2014). Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol.* 21, 64–72. doi: 10.1038/nsmb.2735

Stroud, H., Otero, S., Desvoyes, B., Ramirez-Parra, E., Jacobsen, S. E., and Gutierrez, C. (2012). Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 5370–5375. doi: 10.1073/pnas.1203145109

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* 116, 51–61. doi: 10.1016/S0092-8674(03)01064-X

Takeda, S., Tadele, Z., Hofmann, I., Probst, A. V., Angelis, K. J., Kaya, H., et al. (2004). BRU1, a novel link between responses to DNA damage and epigenetic gene silencing in *Arabidopsis*. *Genes Dev.* 18, 782–793. doi: 10.1101/gad.295404

Talbert, P. B., Masuelli, R., Tyagi, A. P., Comai, L., and Henikoff, S. (2002). Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell* 14, 1053–1066. doi: 10.1105/tpc.010425

Tariq, M., Saze, H., Probst, A. V., Lichota, J., Habu, Y., and Paszkowski, J. (2003). Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8823–8827. doi: 10.1073/pnas.1432939100

Tessadori, F., Chupeau, M.-C., Chupeau, Y., Knip, M., Germann, S., van Driel, R., et al. (2007a). Large-scale dissociation and sequential reassembly of pericentric heterochromatin in dedifferentiated *Arabidopsis* cells. *J. Cell Sci.* 120, 1200–1208. doi: 10.1242/jcs.000026

Tessadori, F., Schulkes, R. K., Van Driel, R., and Fransz, P. (2007b). Light-regulated large-scale reorganization of chromatin during the floral transition in *Arabidopsis*. *Plant J.* 50, 848–857. doi: 10.1111/j.1365-313X.2007.03093.x

Thompson, H. L., Schmidt, R., and Dean, C. (1996). Identification and distribution of seven classes of middle-repetitive DNA in the *Arabidopsis thaliana* genome. *Nucleic Acids Res.* 24, 3017–3022.

van Zanten, M., Koini, M. A., Geyer, R., Liu, Y., Brambilla, V., Bartels, D., et al. (2011). Seed maturation in *Arabidopsis thaliana* is characterized by nuclear size reduction and increased chromatin condensation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20219–20224. doi: 10.1073/pnas.1117726108

Vaquero-Sedas, M. I., and Vega-Palas, M. A. (2013). Differential association of *Arabidopsis* telomeres and centromeres with histone H3 variants. *Sci. Rep.* 3, 1202. doi: 10.1038/srep01202

Volpe, T. A., Kidner, C., Hall, I. M., Teng, G., Grewal, S. I. S., and Martienssen, R. A. (2002). Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297, 1833–1837. doi: 10.1126/science.1074973

Wang, H., Dittmer, T. A., and Richards, E. J. (2013a). *Arabidopsis* CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization. *BMC Plant Biol.* 13:200. doi: 10.1186/1471-2229-13-200

Wang, L.-C., Wu, J.-R., Chang, W.-L., Yeh, C.-H., Ke, Y.-T., Lu, C.-A., et al. (2013b). *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. *J. Exp. Bot.* 64, 1689–1701. doi: 10.1093/jxb/ert030

Wollmann, H., Holec, S., Alden, K., Clarke, N. D., Jacques, P.-E., and Berger, F. (2012). Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the *Arabidopsis* transcriptome. *PLoS Genet.* 8:e1002658. doi: 10.1371/journal.pgen.1002658

Yelagandula, R., Stroud, H., Holec, S., Zhou, K., Feng, S., Zhong, X., et al. (2014). The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in *Arabidopsis*. *Cell* 158, 98–109. doi: 10.1016/j.cell.2014.06.006

Zhang, W., Lee, H.-R., Koo, D.-H., and Jiang, J. (2008). Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. *Plant Cell* 20, 25–34. doi: 10.1105/tpc.107.057083

Zhou, B., Feng, H., and Kato, H. (2013). Structural insights into the histone H1-nucleosome complex. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1222.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

II. Protocole d'hybridation in situ par fluorescence avec des sondes LNA dirigées contre l'ADNr 5S.

Dans le cadre de l'étude de l'ADNr 5S et notamment concernant la question de la localisation et du nombre de clusters, il était nécessaire d'optimiser les outils d'analyse cytologique. En effet, les études précédentes utilisaient une sonde dirigée contre toutes les séquences d'ADNr 5S alors que nous cherchions à marquer spécifiquement un cluster afin de pouvoir le différencier des autres. Pour d'atteindre cet objectif, nous avons décidé de tester l'utilisation de sondes LNA/DNA mixmers qui sont des sondes de grande spécificité (You et al., 2006) jusqu'ici peu utilisées en recherche végétale.

Les sondes LNA/DNA mixmers sont de petites sondes d'environ 20 nucléotides qui pénètrent facilement dans les noyaux. Leurs marquages directs par des fluorochromes ainsi que leurs temps d'hybridation court réduit le temps d'expérimentation.

J'ai optimisé le protocole de FISH avec les sondes LNA/DNA mixmers en particulier la température d'hybridation afin d'obtenir un marquage spécifique pour les *loci* des chromosomes 4 et 5. J'ai réalisé des FISH avec les sondes LNA/DNA mixmers sur des noyaux de cotylédons en interphase ainsi que sur des étalements de chromosomes au stade pachytène.

Les caractéristiques de ces sondes ne nous permettrons pas seulement à l'avenir de différencier les chromosomes, mais aussi de réaliser des FISH sur tissu entier afin de pouvoir observer en 3D l'organisation des gènes d'ARNr 5S.

High-affinity LNA/DNA mixmer probes for detection of chromosome-specific polymorphisms of 5S rDNA repeats in *A. thaliana*

Running title

FISH with LNA/DNA mixmer probes

Authors

Lauriane SIMON[1] and Aline V. PROBST [1]

[1] UMR CNRS 6293 INSERM U1103 Clermont Université, Clermont-Ferrand, France

Abstract

Fluorescence *in situ* hybridization is a standard technique to visualize specific DNA sequences by hybridization with fluorescent probes and, most commonly, relies on DNA probes generated by nick translation. In this chapter, we describe the use of directly labeled LNA/DNA mixmer probes for the rapid detection of repetitive sequences on *Arabidopsis thaliana* nuclei spreads. We further demonstrate that due to the high thermal stability of the heteroduplexes and the resulting elevated binding affinity of LNA/DNA mixmer probes for their target DNA, these probes can be used to discriminate between repetitive sequences differing by only few single nucleotide polymorphisms.

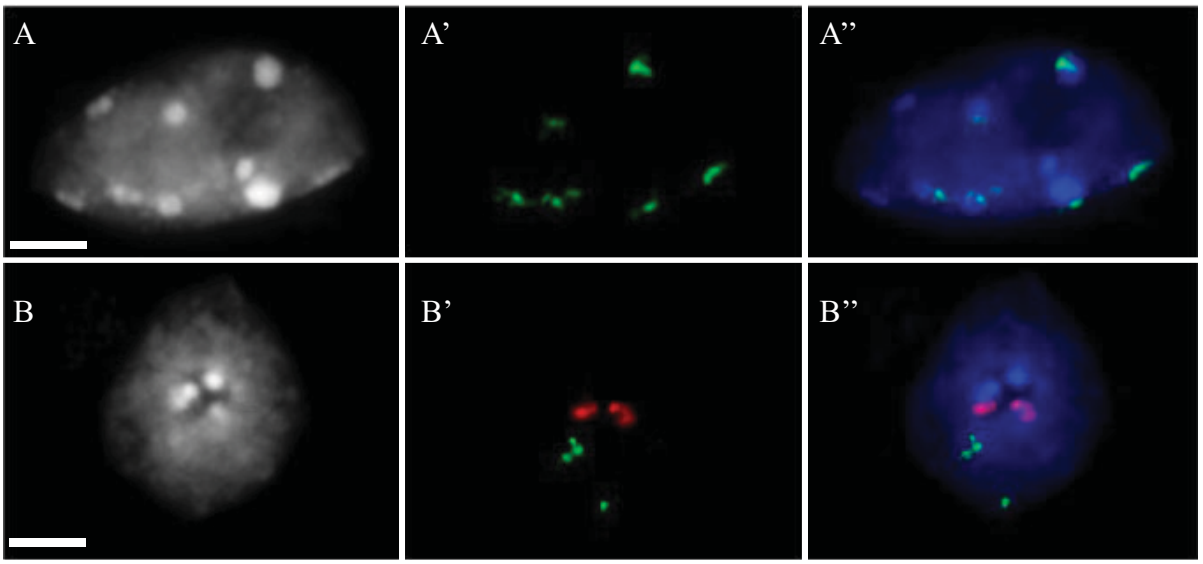
Keywords: *Arabidopsis thaliana*, Fluorescence *in situ* hybridization, LNA/DNA mixmer probes, binding affinity, 5S rDNA, nuclei spreads

1. Introduction

Fluorescence *in situ* hybridization (FISH), developed in the 1980's [1, 2] allows the detection of specific DNA sequences on meiotic chromosome spreads or interphase nuclei [3] and therefore presents a powerful technique both for physical mapping of candidate sequences as well as the study of chromosomal arrangements in a cell-specific manner.

In the model system *Arabidopsis thaliana*, FISH has widely been used to detect repetitive regions and to identify their chromosomal position [4, 5], but has also been applied to visualize whole chromosomes [6] or specific chromosomal domains [7]. FISH allowed both high-resolution physical mapping [7, 8], and investigating the impact of chromatin regulators on nuclear organization of repetitive sequences in interphase [9–11]. Most of the FISH techniques using plant tissues rely on DNA probes generated by nick-translation or random priming-based amplification, using Bacterial Artificial Chromosomes (BACs) as templates for euchromatic regions or cloned tandem repeats as templates for repetitive regions [12]. These DNA probes can be directly labeled by fluorophores or, alternatively, indirectly labeled via the incorporation of Digoxigenin- or Biotin-conjugated nucleotides, subsequently revealed by fluorescently labeled specific antibodies [12]. Immunodetection of Digoxigenin- or Biotin-conjugated nucleotides using several consecutive antibodies amplifies the signal, which is particularly useful for low or single-copy targets. However, the protocol is relatively time consuming and may complicate quantification and cause background staining.

An alternative to longer DNA probes are short oligonucleotide probes and, particularly, Locked nucleic acids (LNA)/DNA mixmers. In these probes some nucleotides are replaced by LNA analogues, in which the ribose ring is "locked" by a methylene bridge between the 2'-O and the 4'-C [13, 14]. This methylene bridge constrains the LNA nucleosides in the ideal conformation for Watson-Crick binding, increasing thermal stability of heteroduplexes and subsequently binding affinity of LNA towards complementary DNA and RNA sequences [15], such that even single nucleotide mismatches might be discriminated [16]. LNA/DNA mixmers have been used to detect short DNA or RNA targets such as miRNAs [17, 18], as well as long non-coding RNAs and repetitive DNA sequences [19] in mammals. Compared to standard probes they might show further advantages due to the resistance to exo-and endonucleases [20, 21], which increases their stability. Furthermore, given that the LNA/DNA mixmers are short oligonucleotides that can be as short as 20 nucleotides, they might



Fluorescence in situ hybridization on cotyledon nuclei using directly labeled LNA/DNA mixmer probes designed to detect all 5S rRNA genes (A, green) or specific for the 5S rRNA genes situated on chromosome 4 (B, red) and chromosome 5 (B, green). A/B, cotyledon nucleus stained with DAPI (gray); A'/B', FISH signals; A''/B'' merged image, DNA in blue. Scale bar: 5 µM.

A. In a diploid nucleus, six 5S rDNA FISH signals are detected corresponding to the *loci* on chromosome 3, 4 and 5. The 5S rDNA repeats partially co-localize with the DAPI-bright chromocenters consisting mainly of centromeric and pericentromeric repetitive elements, but also form loops that emanate from the condensed chromatin domains.

B. The two probes complementary to the chromosome-specific repeat signature specifically recognize two 5S rDNA loci corresponding to chromosome 4 and two corresponding to chromosome 5.

prove advantageous over classical probes in whole-mount preparations. Similar to DNA probes the LNA/DNA mixmers can be coupled to Digoxigenin or Biotin for signal amplification or directly fluorescently labeled, the latter reducing background signals. A drawback is the elevated price of LNA/DNA mixmers compared to traditional methods, however, the amount of the probe provided is sufficient for about 500 hybridization reactions and costs corresponding to antibodies, fluorophores or nick-translation reagents are saved.

Here, we have used directly labeled LNA/DNA mixmer probes to detect the 5S rRNA genes in *Arabidopsis thaliana*, which are highly repeated sequences (more than 1000 copies [22]). The majority of the 5S rRNA genes are clustered into three loci in the pericentromeric regions of chromosomes 3, 4 and 5 in the *A. thaliana* genome of the Colombia-0 (Col-0) accession [23]. Besides the elevated conservation of 5S rRNA gene sequences, the transcription termination sequence just downstream of the transcribed sequence shows a chromosome-specific DNA signature differing by only 6 nucleotides between the 5S rRNA genes of chromosome 4 and 5 [24]. We designed LNA/DNA mixmer probes specific to the transcription termination sequence of 5S rRNA genes of chromosome 4 and 5, optimized hybridization and washing conditions and show that LNA/DNA mixmer probes can indeed be used to discriminate between the two 5S rDNA loci with distinct chromosomal location in nuclei spreads.

This chapter describes a rapid FISH protocol including the preparation of nuclei spreads, hybridization and washing procedures adapted for short LNA/DNA mixmer probes targeting 5S rDNA.

2. Materials

2.1. Fixation of Arabidopsis cotyledons and tissue digestion

1. Carnoy's fixative (*see* **Note 1**): 3:1 Ethanol:Glacial Acetic acid, cold.
2. 2 mL microfuge tubes.
3. Citrate buffer: 10 mM sodium citrate/citric acid in distilled water, pH 4.5 (*see* **Note 2**).
4. Pectolytic enzyme mixture: 0.3 % w/v pectolyase, 0.3 % w/v cytohelicase, 0.3 % w/v cellulase (Sigma) in citrate buffer (*see* **Note 3**).
5. Fine forceps.
6. 6-well plate or small Petri dish (35 x 10 mm).
7. Syringe needles.
8. Incubator at 37°C.

2.2. Preparation of nuclei spreads

1. Cold 60% acetic acid.
2. Syringe needles.
3. Microscope slides (e.g. Star Frost, 76 x 26 mm, Knittel Glass)
4. Hot plate.

5. Cold Carnoy's fixative (*see* **Note 1**): 3:1 Ethanol:Glacial Acetic acid, cold.
6. Coplin Jars.
7. 2X SSC (0.3 M NaCl, 0.03 M sodium citrate, pH 7).
8. Pepsin: 100mL of 100µg/mL Pepsin in 0.01M HCl, prepared freshly.
9. 4 % formaldehyde diluted from liquid 16 % stock (*see* **Note 4, Note 5**) in 1X PBS (10 mM sodium phosphate, pH 7.0, 143 mM NaCl). At least 50 mL in a Coplin Jar is needed.
10. Deionized water.
11. Ethanol series: 70 %, 90 %, and 100 %. At least 50 mL per Coplin Jar is required.
12. Vectashield antifade mounting medium (Vector Laboratories) containing 2 µg/mL DAPI (4',6-diamidino-2-phenylindole).
13. 100 µg/ml DNase free RNase A in 2X SSC.
14. Cover slips 32 x 24 mm.
15. Incubator at 37°C.
16. Moist chamber (*see* **Note 6**).

2.3. Hybridization and Washes

1. Directly labeled LNA/DNA mixmer probes (Exiqon, Denmark) (*see* **Note 7**), 100 µM in 1X PBS:
   - 5S rDNA probe detecting all 5S rRNA genes,
   56FAM_CAAGCACGCTTAACTGCGGAGTTCTGAT"
   - 5S rDNA probe specific for chromosome 4,
   "TEX615_ACCAAAAAAAAAAAAAAAAAAAGAGGGATG" and
   - 5S rDNA probe specific for chromosome 5,
   "56FAM_AAAGGTTAAACATAAAAGAGGGATG".
2. Hybridization buffer (HB50): 50% deionised formamide (Roche), (*see* **Note 8**), 2X SCC, 50 mM Sodium phosphate adjusted to pH 7.
3. "HB50+": 20% dextran sulfate in HB50.
4. Hot plate (*see* **Note 9**).
5. Moist chamber.
6. Incubator at 55°C.
7. Cover slips 32 x 24 mm.
8. 2X SSC.
9. 0.75X SSC.
10. 2X SSC-Tween: 0.1% Tween-20 in 2X SSC.
11. Vectashield antifade mounting medium (Vector Laboratories) containing 2 µg/mL DAPI (4',6-diamidino-2-phenylindole).
12. Fluorescence microscope equipped with optical filters to detect DAPI, FITC/Alexa488/Fam and Texas Red/Cy3 and a digital camera.

3. Methods

3.1. Fixation of Arabidopsis cotyledons and tissue digestion

1. Sample Arabidopsis cotyledons from 10 day-old *in vitro* grown plants (*see* **Note 10**) and place up to 100 cotyledons into a 2 mL microtube containing Carnoy's fixative for fixation overnight at 4°C. Per slide 2 to 3 cotyledons are required.

2. The next morning (*see* **Note 11**), exchange the Carnoy's fixative 2 to 3 times by pipetting until the tissue becomes completely white. The more tissue was sampled, the more Carnoy's fixative exchanges will be needed.
3. Transfer about 10 cotyledons into a 6-well plate with forceps (enough for 3 to 4 slides).
4. Rinse the cotyledons twice with citrate buffer pH 4.5 at room temperature.
5. Remove the citrate buffer by pipetting and add the pectolytic enzyme mixture so that the mixture submerges the cotyledons (6-well plate can be slightly tilted to limit use of pectolytic enzyme mixture). Typically, 50-100 µL are enough to cover 10 cotyledons.
6. Incubate in the incubator at 37°C for about 1.5 h (*see* **Note 12)**.
7. When the tissue is sufficiently digested (*see* **Note 13)**, replace the enzyme mixture by 30 µL cold citrate buffer and proceed directly with preparation of the nuclei spreads.

3.2. Preparation of nuclei spreads

1. Use a needle to lacerate the cotyledons in the 6-well plate until a fine suspension is obtained.
2. Pipette 10 µL of the suspension on a clean slide.
3. Place the slide on a hot plate set to 40°C and add 10 µL of 60 % acetic acid in drops around the suspension. Mix the suspension and acetic acid by gentle circular stirring with a needle (parallel to the glass surface to avoid any contact) for about 1 minute (*see* **Note 14**). This step helps to dissolve the cytoplasm and facilitates nuclear adhesion to the slide. Avoid drying, and if needed, add more acetic acid while stirring.
4. Post-fix the preparation: Add in drops about 500 µL of cold Carnoy's fixative, kept on ice during the procedure, around the suspension and tilt the slide to mix both.
5. Wash the slide 10 minutes in 2X SSC in a Coplin jar at room temperature. All following steps are carried out in Coplin jars. Solutions are exchanged by emptying the Coplin jar, while making sure the slides do not glide out of the jar, and adding fresh solutions into the Coplin jar.
6. Digest the preparation with Pepsin for 3 minutes at 37°C in an incubator (*see* **Note 15**).
7. Wash the slides with 2X SSC for 10 minutes at room temperature.
8. Post-fix the preparation in 4% formaldehyde (*see* **Note 4**) for 10 minutes.
9. Wash the slides with deionized water for 2 minutes.
10. Dehydrate the nuclei by incubating the slides in an ethanol series of 70 %, 90 % and 100 %, 2 minutes each.
11. Take the slides out of the Coplin jar and dry them vertically on tissue paper at room temperature for 10-20 minutes (the slides must be completely dry) (*see* **Note 16**).
12. Some slides can be sacrificed at this step to check the general quality of the preparation. Slides are mounted in Vectashield containing DAPI and monitored under the fluorescence microscope (*see* **Note 17**).
13. Add 100µL of RNase solution on the slide and cover the nuclei preparations with a cover slip. Place the slide in a moist chamber and incubate for 1 h at 37°C. This step removes RNA templates including 5S rRNA, which can cause background staining in the DNA-FISH experiment.
14. Place the slides in a Coplin jar and wash in 2X SSC for 5 minutes at room temperature.
15. Dehydrate the preparation by incubating the slides in a Coplin jar following an ethanol series of 70 %, 90 % and 100 %, 2 minutes each.

16. Take the slides out of the Coplin jar and dry them at room temperature (the slides must be completely dry).

3.3. Hybridization and Washes

1. Place the hot plate under the fume hood and heat it to 80°C.
2. Dilute freshly one or two LNA probes to 0.1 µM in 15 µL HB50 and add 15 µL HB50+.
3. Pipette 30 µL of the hybridization mix per slide onto the nuclei spreads and cover with a cover slip. Avoid forming air bubbles.
4. Place the slides with the hybridization solution on the hot plate for 1 minute. At this step chromosomes and probes are denatured; the time must be carefully monitored.
5. Hybridization: Transfer the slides in a moist chamber and incubate at 55°C for 1 h (*see* **Note 18**).
6. At the same time, place the 2X SSC and 0.75X SSC wash solutions in closed Coplin jars in the 55°C incubator.
7. Remove the coverslips by holding the slide vertically; the coverslip should glide down easily. Transfer in pre-warmed 2X SSC and incubate for 5 minutes at 55°C (*see* **Note 19**). The following steps are carried out in Coplin jars by exchanging the solutions.
8. Wash the slides 5 minutes in 0.75X SSC at 55°C (*see* **Note 20**).
9. Wash the slides 5 minutes in 2X SSC at 55°C.
10. Wash the slides 5 minutes in 2X SSC-Tween at room temperature.
11. Rinse the slide rapidly with deionized water.
12. Air-dry the slides.
13. Mount the slides in a drop of Vectashield containing DAPI (*see* **Note 21**).
14. Observe the slides under the fluorescence microscope (*see* **Note 22**).

4. Notes

1. Carnoy's fixative should be prepared freshly on ice.
2. To prepare the citrate buffer 10X stock solution, add 0.1 M citric acid monohydrate into 0.1 M trisodium citrate dehydrate until pH 4.5 is reached. The 10X stock can be stored at 4°C and should be diluted into double distilled water prior to use.
3. We generally use the following products (C8274, C1794, P5936; all from Sigma) and prepare a stock solution containing 1% of each enzyme in citrate buffer, which is stored in small aliquots at –20°C. Variations between batches may exist and digestion times may need to be adjusted accordingly for each new enzyme mixture batch.
4. Formaldehyde dilutions should be freshly prepared. Once opened, formaldehyde bottles should not be stored for longer periods of time to avoid concentration changes. We advice to use Formaldehyde supplied in small glass vials containing 10 mL of 16 % Formaldehyde (e.g. ThermoFisher 28908) and to discard left-over dilutions.
5. Formaldehyde and Formamide are toxic products in case of contact or inhalation. Dilutions and post-fixation steps as well as denaturation should be carried out under the fume hood and using adequate protective equipment.

6. We typically place the slides horizontally into plastic slide storage boxes with a lid and a tissue imbibed with 2X SSC at the bottom. The slides should not be in contact with the imbibed tissue.
7. We have successfully used LNA/DNA mixmer probes between 20 and 25 nucleotides long. The best positions of the LNA analogues in the probe sequence are determined by the provider [15, 16]. The probes are delivered in dry format and, upon arrival, should be diluted in 1X PBS to a stock solution of 100 µM. The stock solutions are stored in aliquots at -20°C. While we present here results for the 5S rDNA probes, other repetitive sequences such as 180bp centromeric repeats are ideal targets for LNA/DNA mixmer probes.
8. To avoid pH changes, deionised formamide solutions should be stored in small aliquots at -20°C.
9. The temperature on the hot plate should be monitored and it is important that the surface is perfectly planar to allow uniform transfer of heat onto the slide.
10. Other tissues and other developmental stages such as leaves or flower buds can be used, but incubation times with the cell wall digestion mix should be adapted. See also Note 11. Flower buds contain both interphase nuclei, metaphase spreads as well the pollen mother cells at the pachytene stage, the latter having the advantage that chromosomes are synapsed and about 25 times longer than mitotic chromosomes [12].
11. The whole protocol can be carried out in one day, since the hybridization time is short (1 hour) compared to standard DNA probes, which require an overnight hybridization step.
12. Depending on the tissue, the digestion time needs to be adjusted. For inflorescences the digestion time should be increased to 3 h.
13. Check if the tissue is sufficiently digested by gently squashing one cotyledon, which should release a white cloud corresponding to cells in suspension – if only fragments are released, incubate longer.
14. This is one of the most critical steps: acetic acid treatment is needed to clear and dissolve cytoplasmic components. Insufficient clearing creates clumps and layers of cellular fragments that will cause background during hybridization and DNA staining. The nuclei should appear smooth with a regular contour. Excess of treatment leads to nuclei with irregular contours and clumps in DNA staining.
15. This step is optional, but helps to remove remaining cytoplasmic debris and improves access to the DNA probe. It is important to monitor precisely the incubation time and temperature. Possibly, the incubation time can be reduced when damaged nuclei are observed. Excess of treatment damages the chromatin and can lead to poor hybridization signals.
16. Slides can be stored at this step in a closed box at 4°C.
17. The nuclei should appear isolated and not aggregated in clumps. No cytoplasm should remain around the nuclei. DAPI staining should reveal nuclei with smooth and regular contours, allowing to clearly distinguishing euchromatin (lightly stained) and heterochromatic regions (strongly stained). To avoid clumps of cells, increase digestion time and squash cotyledons more extensively.
    If excessive amounts of cytoplasm remain, the tissue digestion time can be adjusted and the density of cells reduced by using less starting material. Furthermore, during clearing,

the amount of acetic acid can be increased and/or the stirring time prolonged (*see* **Note 14**). In addition, the time of pepsin treatment can be adjusted.

18. The temperature and the time of hybridization must be adjusted depending on the probe. When hybridization is performed with the 5S rDNA probe that detects all 5S rRNA genes, hybridization can be carried out at lower temperatures such as 37°C. To achieve labeling specificity with the probes designed to hybridize exclusively to the 5S rDNA loci situated on chromosome 4 or 5, hybridization should be performed at 55°C.
19. It is important that the wash solutions have reached the required temperature.
20. If non-specific binding of the probes is observed, the stringency of washing conditions can be increased: by lowering SSC concentrations, by adding Formamide to the wash buffers or by increasing the washing temperatures. Hybridization temperature can also be increased (see Note 18).
21. Coverslips can be sealed with nail polish and stored at 4°C for several weeks.
22. For image acquisition, our typical microscope setup comprises a Leica SIM microscope, equipped with optical filters for DAPI, FITC and TexasRed, a black and white digital CMOS camera (Hamamatsu, ORCA-Flash 4.0 V2) that produces 16-bit images, and Metamorph imaging software. To image the DNA stained by DAPI, the typical exposure time is 10 msec, for FISH signals in green and red channels, 100 msec is currently used. The appropriate exposure time should be adjusted for each channel. To create overlays of the black and white images acquired in the blue, green or red channel, the ImageJ ‘merge channels’ option is used. Alternatively, images can be merged with Photoshop; however, care must be taken to monitor possible xy chromatic shifts.

5. References

Bauman, J., Wiegant, J., Borst, P., and van Duijn, P. (1980). A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. *Exp. Cell Res.* 128, 485–90.

Campell, B. R., Song, Y., Posch, T. E., Cullis, C. A., and Town, C. D. (1992). Sequence and organization of 5S ribosomal RNA-encoding genes of Arabidopsis thaliana. *Gene* 112, 225–228. doi:10.1016/0378-1119(92)90380-8.

Cloix, C., Tutois, S., Yukawa, Y., Mathieu, O., Cuvillier, C., Espagnol, M.-C., et al. (2002). Analysis of the 5S RNA Pool in Arabidopsis thaliana: RNAs Are Heterogeneous and Only Two of the Genomic 5S Loci Produce Mature 5S RNA. *Genome Res.* 12, 132–144. doi:10.1101/gr.181301.

Crinelli, R. (2002). Design and characterization of decoy oligonucleotides containing locked nucleic acids. *Nucleic Acids Res.* 30, 2435–2443. doi:10.1093/nar/30.11.2435.

Fransz, P. F., Armstrong, S., Alonso-Blanco, C., Fischer, T. C., Torres-Ruiz, R. A., and Jones, G. (1998). Cytogenetics for the model system Arabidopsis thaliana. *Plant J.* 13, 867–76.

Fransz, P. F., Armstrong, S., de Jong, J. H., Parnell, L. D., van Drunen, C., Dean, C., et al. (2000). Integrated cytogenetic map of chromosome arm 4S of A. thaliana: structural organization of heterochromatic knob and centromere region. *Cell* 100, 367–76.

Fransz, P., de Jong, J. H., Lysak, M., Castiglione, M. R., and Schubert, I. (2002). Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14584–14589.

Frieden, M., Hansen, H. F., and Koch, T. (2003). Nuclease stability of LNA oligonucleotides and LNA-DNA chimeras. *Nucleosides. Nucleotides Nucleic Acids* 22, 1041–3. doi:10.1081/NCN-120022731.

Koshkin, A. A., Singh, S. K., Nielsen, P., Rajwanshi, V. K., Kumar, R., Meldgaard, M., et al. (1998). LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron* 54, 3607–3630. doi:10.1016/S0040-4020(98)00094-5.

Langer-Safer, P. R., Levine, M., and Ward, D. C. (1982). Immunological method for mapping genes on Drosophila polytene chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 79, 4381–5. doi:10.1073/pnas.79.14.4381.

Levsky, J. M., and Singer, R. H. (2003). Fluorescence in situ hybridization: past, present and future. *J Cell Sci* 116, 2833–2838. doi:10.1242/jcs.00633.

Lysak, M., Fransz, P., and Schubert, I. (2006). Cytogenetic analyses of Arabidopsis. *Methods Mol Biol* 323, 173–186. doi:10.1385/1-59745-003-0:173.

Maluszynska, J., and Heslop-Harrison, J. (1993). Molecular cytogenetics of the genus Arabidopsis: in situ localization of rDNA sites, chromosome numbers and diversity in centromeric heterochromatin. *Ann. Bot.* 71, 479–484. doi:10.1006/anbo.1993.1063.

Maluszynska, J., and Heslop-Harrison, J. S. (1991). Localization of tandemly repeated DNA sequences in Arabidopsis thaliana. *Plant J.* 1, 159–166. doi:10.1111/j.1365-313X.1991.00159.x.

Mathieu, O., Jasencakova, Z., Vaillant, I., Gendrel, A.-V., Colot, V., Schubert, I., et al. (2003). Changes in 5S rDNA Chromatin Organization and Transcription during Heterochromatin Establishment in Arabidopsis. *Plant Cell* 15, 2929–2939. doi:10.1105/tpc.017467.

Mittelsten Scheid, O., Probst, A. V, Afsar, K., and Paszkowski, J. (2002). Two regulatory levels of transcriptional gene silencing in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 99, 13659–62. doi:10.1073/pnas.202380499.

Obika, S., Nanbu, D., Hari, Y., Morio, K., In, Y., Ishida, T., et al. (1997). Synthesis of 2′-O,4′-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C3, -endo

sugar puckering. *Tetrahedron Lett.* 38, 8735–8738. doi:10.1016/S0040-4039(97)10322-7.

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M. A., et al. (2004). Chromosome territory arrangement and homologous pairing in nuclei of Arabidopsis thaliana are predominantly random except for NOR-bearing chromosomes. *Chromosoma* 113, 258–269.

Probst, A. V., Fransz, P. F., Paszkowski, J., and Mittelsten Scheid, O. (2003). Two means of transcriptional reactivation within heterochromatin. *Plant J.* 33, 743–749.

Probst, A. V, Okamoto, I., Casanova, M., El Marjou, F., Le Baccon, P., and Almouzni, G. (2010). A strand-specific burst in transcription of pericentric satellites is required for chromocenter formation and early mouse development. *Dev. Cell* 19, 625–38. doi:10.1016/j.devcel.2010.09.002.

Silahtaroglu, A. N., Tommerup, N., and Vissing, H. (2003). FISHing with locked nucleic acids (LNA): Evaluation of different LNA/DNA mixmers. *Mol. Cell. Probes* 17, 165–169. doi:10.1016/S0890-8508(03)00048-3.

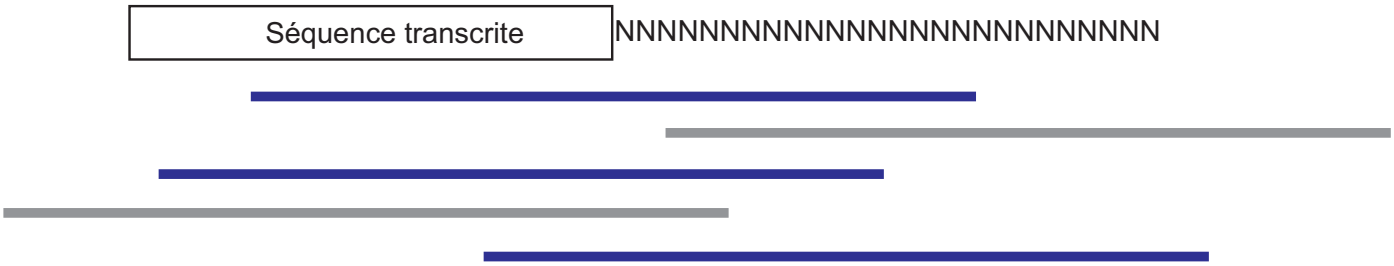
Søe, M. J., Møller, T., Dufva, M., and Holmstrøm, K. (2011). A sensitive alternative for microRNA in situ hybridizations using probes of 2'-O-methyl RNA + LNA. *J. Histochem. Cytochem.* 59, 661–72. doi:10.1369/0022155411409411.

Song, R., Ro, S., and Yan, W. (2010). In situ hybridization detection of microRNAs. *Methods Mol. Biol.* 629, 287–94. doi:10.1007/978-1-60761-657-3_18.

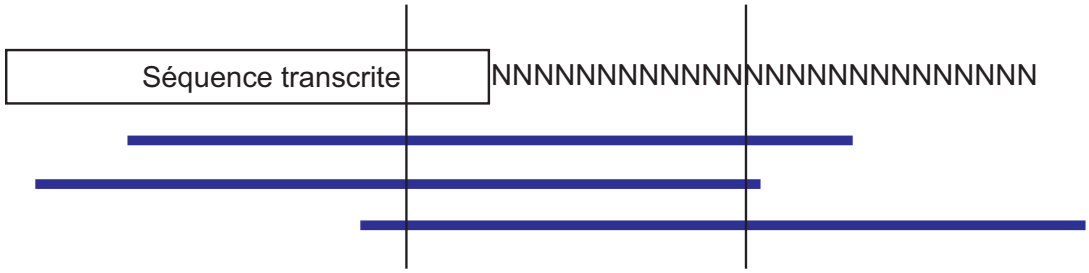
You, Y., Moreira, B. G., Behlke, M. A., and Owczarzy, R. (2006). Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res.* 34, 1–11. doi:10.1093/nar/gkl175.

Annexe III

1) Aligner les reads sur la séquence transcrite consensus prolongée de 30 nucléotides en 3'



2) Garder uniquement les reads couvrant la fin de la séquence transcrite +15 bases en aval



3) Extraire les 15 premières bases des T-stretchs ainsi que le nombre de reads contenant ces T-stretchs

TTTTATGTTTAACCT	3126
ATTTTTTTTTTTTTT	847
TTTTTTTTTTTTTTT	757
TTTTTTTGTTTCTTT	310
TTTTATGTTTAACCT	255
CTTTTTTTTTTTTTT	118
TTTTTTTTTTTTTTT	100
TTTTATGTTTAACAT	97

III. Identification des T-stretch

Tristan Dubos (M2R Analyses et Modélisation des données ; 2014-15) a mis au point un pipeline bio-informatique permettant d'extraire tous les T-stretchs présents dans des données de séquençage à haut débit de type Hi-Seq Illumina avec des reads de 50 pb, et de les comptabiliser.

La première étape consiste à mapper les reads avec **bwa ALN** (Li and Durbin, 2009) en utilisant comme séquence de référence la séquence transcrite de l'ADNr 5S de 120 bp allongée de 30 N en 3'. Seuls seront pris en compte, les reads qui mappent sur les 20 dernières bases de la séquence transcrite avec moins de 5 mismatchs.

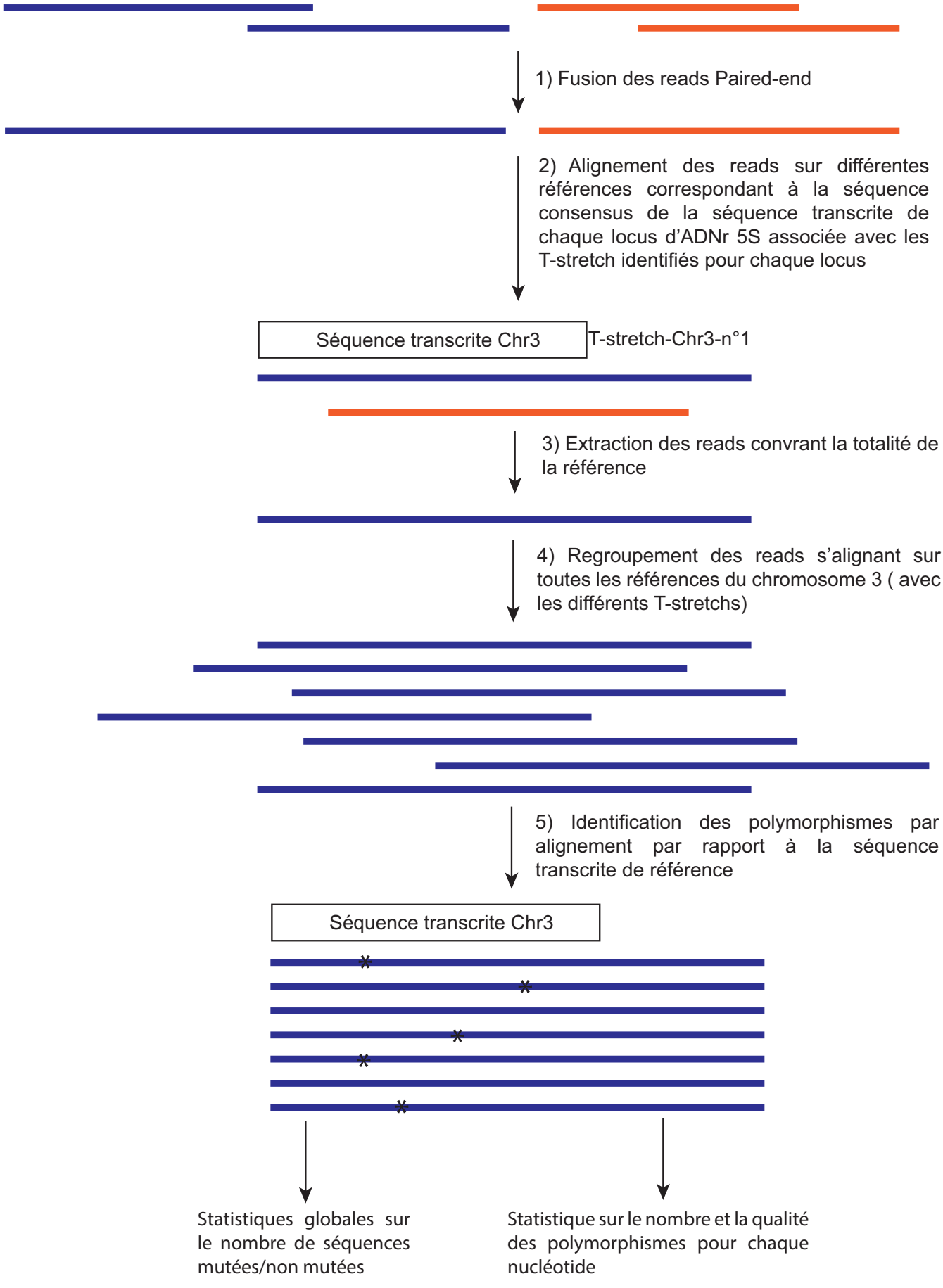
Par la suite, les 15 premières bases de la séquence T-stretch sont extraites pour chaque read et compilées dans un fichier contenant la séquence de ces 15 bases et le nombre de reads portant cette signature.

Par la suite, chaque T-stretch est attribué à un chromosome grâce à un tri manuel.

Référence :

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–60. doi:10.1093/bioinformatics/btp324.

Annexe IV


1) Fusion des reads Paired-end
2) Alignement des reads sur différentes références correspondant à la séquence consensus de la séquence transcrite de chaque locus d'ADNr 5S associée avec les T-stretch identifiés pour chaque locus
Séquence transcrite Chr3
T-stretch-Chr3-n°1
3) Extraction des reads convrant la totalité de la référence
4) Regroupement des reads s'alignant sur toutes les références du chromosome 3 (avec les différents T-stretchs)
5) Identification des polymorphismes par alignement par rapport à la séquence transcrite de référence
Séquence transcrite Chr3
Statistiques globales sur le nombre de séquences mutées/non mutées
Statistique sur le nombre et la qualité des polymorphismes pour chaque nucléotide

IV. Identification des polymorphismes de la séquence transcrite

Damien Laubert (M2R Analyses et Modélisation des données; 2015-16) a mis au point un pipeline bio-informatique permettant d'extraire tous les polymorphismes présents dans des données de séquençage à haut débit Illumina de type Miseq paired-end avec des reads de 250 bp et de les comptabiliser.

Les reads pris en compte doivent recouvrir la totalité de la séquence transcrite et du T-stretch.La première étape consiste à créer des « reads longs ». Pour se faire, le read long est créé à partir des 2 reads issus d'un séquençage paired-end par le logiciel FLASH (Magoč and Salzberg, 2011).

Les reads longs sont ensuite mappés avec **bwa MEM** (Li, 2013) sur différentes références constituées de la séquence transcrite de l'ADNr 5S allongée de chaque T-stretch préalablement identifié (Annexe III). Seuls les reads recouvrant la totalité de la séquence transcrite sont conservés.

Les reads ayant mappés sur les références de chaque chromosome sont regroupés ensembles et alignés. Par la suite, le pipeline comptabilise pour chaque chromosome, le nombre et la qualité des polymorphismes sur chaque nucléotide de la séquence transcrite.

De plus, le nombre de reads avec aucun (ARNr 5S majoritaire), un et deux (ARNr minoritaire) ou plus de deux polymorphismes (ARNr 5S non transcrit) est comptabilisé pour chaque cluster.

On peut aussi utiliser ce pipeline sur la séquence totale du gène d'ARNr 5S et ainsi obtenir de nouvelles séquences de référence d'une unité complète en modifiant la séquence de référence précédente dès qu'un polymorphisme est présent dans plus de 50% reads issus d'un cluster.

Références :

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 3.

Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–63. doi:10.1093/bioinformatics/btr507

V. Rôle de du complexe CAF-1 dans la formation des chromocentres

La formation et l'organisation de l'hétérochromatine est importante pour la stabilité du génome et la répression des éléments transposables. Comme décrit précédemment, chez *Arabidopsis*, les régions hétérochromatiques s'associent dans le noyau en structures denses, fortement colorés au DAPI : les chromocentres. Les mécanismes impliqués dans la formation de ces chromocentres et ce notamment pendant les premiers jours suivant la germination pendant lesquels les pré-chromocentres sont consolidés en chromocentres matures restent peu étudiées. Nous nous sommes donc demandé quel était le rôle des modifications post-traductionnelles des histones et des variants d'histones dans l'organisation de la chromatine et la formation des chromocentres.

Nous avons montré, par des analyses de ChIP-qPCR sur des cotylédons âgés de 2 jours après germination, stade auquel seuls des pré-chromocentres sont présents, et de 5 jours, stade auquel les chromocentres matures sont formés, que les séquences centromériques et péricentromériques s'enrichissent en marque H3K9me2. De plus, l'analyse de mutants de protéines impliquées dans le dépôt des modifications post-traductionnelles des histones nous a permis de montrer que la déposition correcte des marques H3K9me2 et H3K27me1 est importante pour la formation des chromocentres et la compaction des séquences répétées. Nous avons aussi observé un enrichissement en histone canonique H3.1 au niveau des séquences répétées au fur et à mesure de la formation des chromocentres. Cette dynamique de dépôt d'histone H3.1 au niveau des régions hétérochromatiques requiert l'intervention du complexe de chaperonne d'histone, CAF-1. En effet, la perte de CAF-1 mais pas celle du complexe HIR altère le dépôt d'H3.1 et compromet la formation des chromocentres.

J'ai participé à l'élaboration de ce manuscrit par des expériences de ChIP-qPCR et des analyses d'enrichissement d'H3.1 au niveau de chromocentres par des approches d'immunofluorescence.

Role of Chromatin Assembly Factor 1 mediated H3.1 deposition in chromocenter formation

Author affiliation:

Matthias Benoit, Lauriane Simon, Céline Duc, Sylviane Cotterell, Axel Poulet, Samuel Le Goff, Christophe Tatout and Aline V. Probst

Génétique, Reproduction et Développement, UMR CNRS 6293, Clermont Université, INSERM U1103, 24 Avenue des Landais, BP 80026, 63171 Aubière Cedex, France

Summary:

Using chromocenter formation during post-germination development in cotyledons as a paradigm, we show that spatial organization of centromeric and pericentromeric heterochromatin is dynamic and clustering into chromocenters reinforces transcriptional silencing at pericentromeric repeats. H3K27me1 and H3K9me2 are already enriched at centromeric and pericentromeric heterochromatin compared to euchromatin despite the absence of conspicuous chromocenter structures. H3K9me2 levels further increase during chromocenter formation and the H3K9 dimethyltransferases SUVH4, SUVH5 and SUVH6, as well as the H3K27 monomethyltransferases ATXR5 and ATXR6 are required for proper kinetics of chromocenter formation. In parallel, we show that the clustering of heterochromatic repeats in chromocenters requires the Chromatin Assembly Factor 1 (CAF-1) histone chaperone, which is responsible for H3.1 enrichment specifically at heterochromatic repeats. We conclude that chromocenter formation is driven by local enrichments in repressive histone marks and the canonical H3.1, arguing for a critical role of replication-coupled H3.1 nucleosome assembly and the deposition of repressive histone modifications in this developmental process.

Abstract

Chromocenter organization compartmentalizes heterochromatin from the remaining genome; however, the mechanisms involved in chromocenter formation are not understood. By characterizing heterochromatin organization during *Arabidopsis thaliana* post-germination development, we show that enrichment in the repressive marks H3K9me2 and H3K27me1 precedes chromocenter formation, and that H3K9me2 and H3K27me1 methyltransferases are required for clustering of repetitive elements. Furthermore, levels of the canonical histone H3.1 increase specifically at heterochromatic repeats during repeat clustering. This H3.1 enrichment and chromocenter formation require the Chromatin Assembly Factor 1 (CAF-1) chaperone complex, demonstrating a critical function for CAF-1-mediated histone deposition and concomitant modification of histone tails in chromocenter formation.

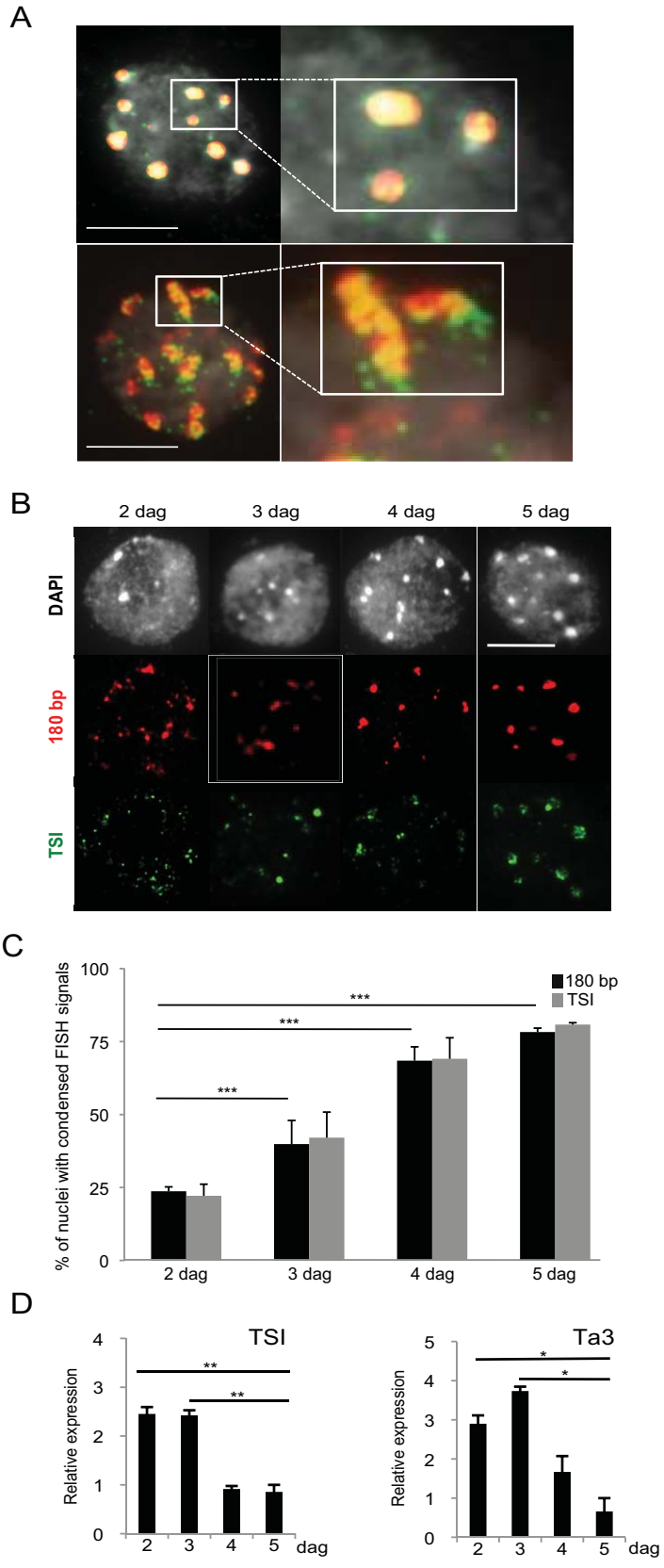


Fig. 1: Progressive organization and enhanced silencing of centromeric and pericentromeric repeats during chromocenters formation.

(A-B) Representative nuclei after DNA FISH with probes for 180 bp (red) and TSI (green) sequences and counterstained with DAPI (grey). Scale bars, 5 µm. (A) Example of a clustered (top) and dispersed (bottom) organization. (B) Nuclei from cotyledons aged 2, 3, 4 or 5 days after germination (dag). (C) Percentage of nuclei (± SEM from 2 biological replicates) with complete clustering of either 180 bp (black) or TSI sequences (grey) in chromocenter structures, n>200. *p<0.05; **p<0.01; ***p<0.005; χ2 test. (D) RT-qPCR analysis of TSI (left) and Ta3 (right) transcripts in WT cotyledons. Histograms show mean fold change (± SEM obtained for at least 2 biological replicates relative to transcript levels at 5 dag (set to 1) after normalization to At2g28390. *p<0.05; **p<0.01; Student's t-test.

Introduction

In eukaryotes, the genetic information is organized in chromatin. Different arrangements of the array of nucleosomes, the basic subunits of chromatin, result in distinct higher-order chromatin structures such as condensed heterochromatin domains. Heterochromatin comprises mainly transposable elements and other repetitive sequences and is critical for genome integrity as it prevents illegitimate recombination and transcription of these genetic elements. In certain yeasts, animals and plants, heterochromatin domains cluster in cytologically distinct chromocenters. Chromocenter organization is thought to contribute to heterochromatin function as it physically separates the repetitive elements from the rest of the genome and may so favor the concentration of silencing factors (Almouzni and Probst, 2011). Indeed, heterochromatic repeats in chromocenters are enriched in epigenetic marks that repress transcription, including DNA methylation and specific post-translational modifications (PTMs) of histones such as hypoacetylation and methylation of H3K9me2/3 and H3K27me1 (Lehnertz et al., 2003; Martens et al., 2005; Mathieu et al., 2005; Probst et al., 2003). These repeats further show high nucleosomal density (Chodavarapu et al., 2010) and low DNase I accessibility (Shu et al., 2012). In addition to histone PTMs, the incorporation of various histone variants can change the composition of the nucleosome, which profoundly affects stability and ultimately higher-order organization of the chromatin fiber (Volle and Dalal, 2014). While the H3 variant H3.3, found at actively transcribed genes, creates more unstable nucleosomes contributing to accessible chromatin structures (Goldberg et al., 2010; Jin and Felsenfeld, 2007; Mito et al., 2005), the canonical H3.1 is found in transcriptionally inactive heterochromatin (Goldberg et al., 2010; Stroud et al., 2012; Wollmann et al., 2012).

The organization of repetitive sequences in chromocenters, their number and distribution as well as the chromosomes involved in an individual chromocenter can vary in different cell types and during differentiation (Fransz et al., 2006; Mayer et al., 2005), suggesting that the nuclear organization of heterochromatin has a role beyond transcriptional control of the underlying repeat sequences. Indeed, in Arabidopsis, chromocenters function as organizing centers for gene-rich euchromatic loops within a chromosome territory (Feng et al., 2014; Fransz et al., 2002b; Grob et al., 2014). Interestingly, chromocenter organization is not static, but can be profoundly affected in response to developmental or environmental cues, which can induce important changes in gene expression patterns (Benoit et al., 2013; Pecinka et al., 2010; Tessadori et al., 2007a). A transient decondensation of chromocenter structures takes place during imbibition of the seed and immediately after germination, centromeric and pericentromeric repeats are dispersed in the nucleus (van Zanten et al., 2011). Within a few days, chromocenters are reformed (Benoit et al., 2013; Douet et al., 2008; Mathieu et al., 2003a). However, the molecular features of centromeric and pericentromeric repeats at the different states of higher-order organization and the mechanisms that drive chromocenter formation are not understood so far.

Here we show that centromeric and pericentromeric repeats change their spatial organization to cluster into chromocenters in a precise temporal arrangement that coincides with a reinforcement of

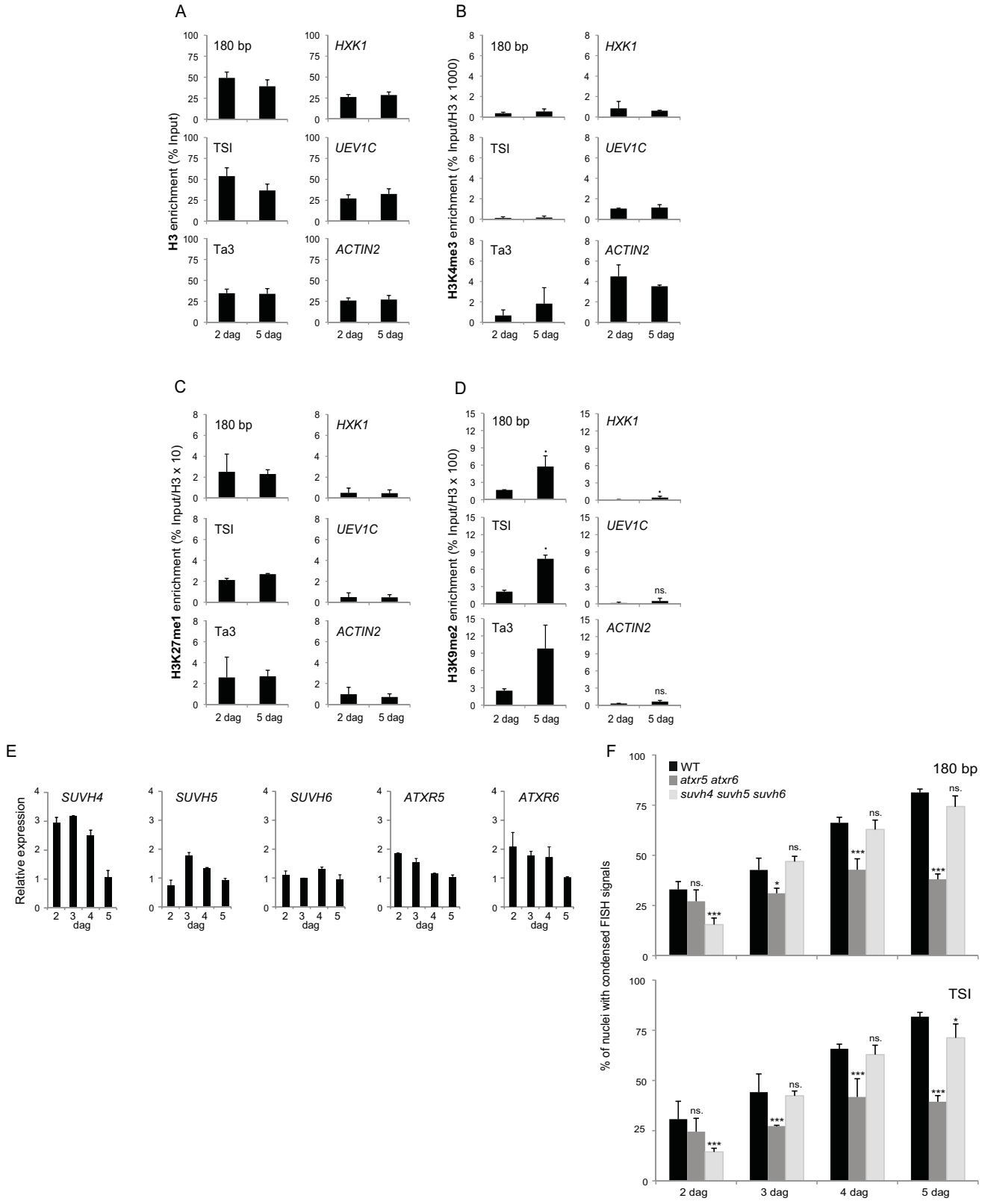


Fig. 2: Repressive marks are enriched at centromeric and pericentromeric sequences prior to chromocenter organization and required for repeat clustering

Histone H3 occupancy relative to input (A) and enrichment of H3K4me3 (B), H3K27me1 (C) and H3K9me2 (D) normalized to H3-levels in WT cotyledons at 2 or 5 dag as assessed by ChIP-qPCR at heterochromatic elements (180 bp, TSI and the Ta3 retrotransposon) and euchromatic loci (HXK1, UEV1C and ACTIN2). Histograms show mean percentage ± SEM for two independent PCR amplifications of three biological replicates. ns. p>0.1; •p<0.1; *p<0.05; Student's t-test. (E) RT-qPCR analysis of SUVH4, SUVH5, SUVH6, ATXR5 and ATXR6 transcripts in WT cotyledons aged 2, 3, 4 or 5 dag. Histograms show mean transcript levels relative to 5 dag (set to 1) after normalization to At2g28390 ± SEM obtained for two independent PCR amplifications of three biological replicates. (F) Percentage of nuclei (± SEM from 2 biological replicates) with clustered 180 bp (top) and TSI (bottom) sequences in WT, *atxr5 atxr6* and *suvh4 suvh5 suvh6* mutants, n>200. *p< 0.05; **p< 0.01; ***p< 0.005; χ2 test.

transcriptional repression of pericentromeric repeats. Before conspicuous chromocenter structures are established, repressive histone marks are already enriched at centromeric and pericentromeric repeats compared to euchromatin. H3K9me2 levels further increase during post-germination development and the histone methyltransferases responsible for H3K9 di- or H3K27 monomethylation are required for proper kinetics of repeat clustering. Finally, we show that chromocenter formation correlates with an increase in canonical H3.1 specifically at heterochromatic regions and that H3.1 loading and chromocenter formation require the CAF-1 histone chaperone. We conclude that chromocenter formation is associated with local enrichments in canonical histones and their modifications, suggesting that replication-coupled H3.1 deposition and installation of repressive PTMs play a critical role in this process.

Results

Progressive formation of chromocenters during post-germination development is associated with transcriptional repression

To quantify the clustering of heterochromatic repeats into chromocenters during early post-germination development, we determined the percentage of nuclei with fully clustered or dispersed centromeric (180 bp) and pericentromeric (Transcriptionally Silent Information (TSI) (Probst et al., 2003)) repeats that are present on all chromosomes (Fig. 1). At 2 days after germination (dag) the majority of cotyledon nuclei exhibit small pre-chromocenters visible by DAPI-staining and only about 22% of the nuclei reveal completely clustered repeats compared to ~80% at 5 dag (Fig. 1B-C). We find that centromeric and pericentromeric sequences cluster with a precise temporal arrangement and with similar kinetics into chromocenters during this developmental time window (Fig. 1C). Centromeric 180 bp repeats, TSI as well as most pericentric transposable elements are transcriptionally silenced (May et al., 2005; Probst et al., 2003; Steimer et al., 2000). To test whether chromocenter formation contributes to transcriptional repression of repetitive elements, we quantified transcript levels of TSI repeats and the Ta3 retrotransposon located in the pericentromeric region by RT-qPCR from total RNAs isolated from dissected cotyledons. We reproducibly observed higher levels of TSI and Ta3 transcripts in 2 dag-old cotyledons, when these genomic regions are dispersed in the large majority of nuclei (Fig. 1D), compared to 5 dag. We conclude that heterochromatic repeats progressively cluster into chromocenters betweeen 2 to 5 dag and that chromocenter formation correlates with reduced levels of TSI and Ta3 transcripts.

H3K9me2 and H3K27me1 enrichments precede chromocenter formation and their defective setting affects kinetics of chromocenter formation

To gain insight into possible changes of local chromatin features during chromocenter formation at these repetitive elements and to examine whether specific PTMs could be required for repeat clustering we used Chromatin Immunoprecipitation (ChIP) and FISH (Fig. 2). We first dissected cotyledons from seedlings aged 2 or 5 dag and precipitated chromatin with antibodies directed against H3, H3K4me3, a histone mark associated with transcriptional activity (Roudier et al., 2011a) and two repressive marks H3K9me2 and H3K27me1 (Mathieu et al., 2005; Probst et al., 2003). Nucleosome occupancy, as reported by the H3 levels, is higher at 180 bp and TSI repeats compared

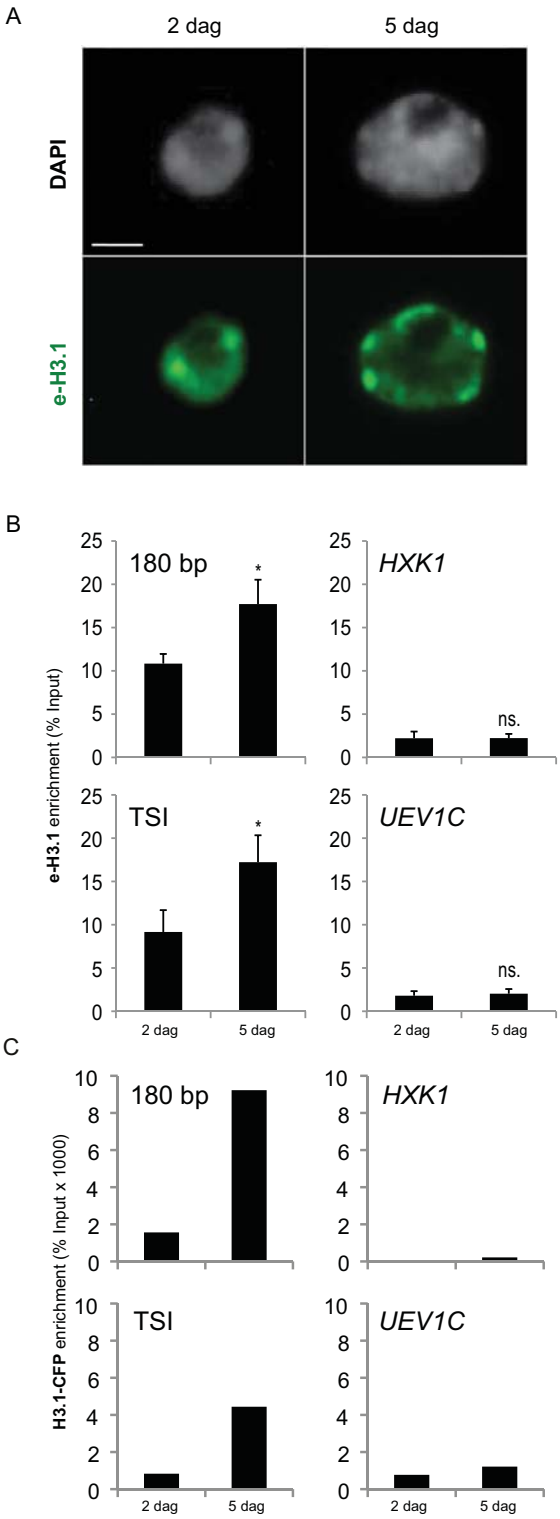


Figure 3: H3.1 levels at heterochromatin increase during post-germination growth.

(A) Subnuclear localization of e-H3.1 revealed by immunostaining (green) in nuclei of cotyledons aged 2 or 5 dag. DNA is counterstained with DAPI (grey). Scale bar, 5 µm. (B) e-H3.1 enrichment determined by FLAG-ChIP qPCR relative to input in cotyledons from WT plants aged 2 or 5 dag at heterochromatic elements (180 bp, TSI) and two euchromatic loci (HXK1, UEV1C). Histograms show mean percentage ± SEM of FLAG-immunoprecipitation relative to input for 3 biological replicates. *p<0.05; Student's t-test. (C) Enrichment of H3.1-CFP as determined by ChIP at 180 bp, TSI, HXK1 and UEV1C. Histograms show mean percentage of two independent PCR amplifications.

to euchromatic loci, but does not seem to be modified during the process of chromocenter formation (Fig. 2A). While H3K4me3 levels are low in heterochromatin throughout the developmental time window (Fig. 2B), the repressive marks H3K27me1 and H3K9me2 are 6- or 10-fold, respectively, enriched at heterochromatin compared to the three euchromatic loci (*HXK1, UEV1C* and *ACTIN2*) (Fig. 2C-D) and this as early as 2 dag when chromocenters are not yet fully formed. While H3K27me1 levels remain rather stable at all targets (Fig. 2C), H3K9me2 levels increase between 2 and 5 dag (Fig. 2D). We then examined to what extent the appropriate setting of these repressive marks affects chromocenter formation. The SET domain proteins SU(VAR)3-9 HOMOLOG 4/KRYPTONITE (KYP), SUVH5 and SUVH6 (Ebbs and Bender, 2006; Jackson et al., 2002) dimethylate H3K9 whereas H3K27me1 at heterochromatin is set by the histone methyltransferases ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 (ATXR5) and ATXR6 (Jacob et al., 2014; Jacob and Feng, 2009). The genes encoding these methyltransferases are expressed during this developmental window (Fig. 2E). To evaluate the importance of H3K27me1 and H2K9me2 marks for chromocenter formation, we scored the percentage of nuclei with clustered 180 bp and TSI repeats in *suvh4 suvh5 suvh6* and *atxr5 atxr6* mutant cotyledons. In *suvh4 suvh5 suvh6* mutants, the percentage of nuclei with complete clustering of centromeric and pericentromeric repeats is significantly lower at 2 dag, but then reaches levels similar to WT with the exception of the pericentromeric TSI repeats, for which less nuclei showed complete clustering at 5 dag compared to WT (Fig. 2F). In contrast, in *atxr5 atxr6* cotyledons, only 40% of the nuclei complete chromocenter formation at 4 and 5 dag, compared to 70% and 80%, respectively, in the WT (Fig. 2F). Deficient H3K27 monomethylation in absence of ATXR5 and ATXR6 therefore severely impairs clustering of 180 bp and TSI repeats into chromocenter structures.

Taken together, centromeric and pericentromeric chromatin domains already show enrichment in repressive histone marks compared to active genes as early as 2 dag, despite the dispersion of the repetitive elements. H3K9me2 levels further increase at heterochromatic repeats between 2 and 5 dag, and the appropriate setting of both H3K9me2 and H3K27me1 affects chromocenter formation.

H3.1 becomes enriched at heterochromatin after germination

H3K9me2 and H3K27me1 are predominantly found enriched at the canonical histone H3.1 (Jacob et al., 2014; Johnson et al., 2004; Loyola et al., 2006). We were therefore interested to investigate whether there is a functional link between H3.1 enrichment and clustering of repetitive sequences into chromocenters (Fig. 3). Since no antibodies able to distinguish the Arabidopsis canonical H3.1 from its variant counterparts are available, we generated a transgenic line expressing *HTR9*-encoded H3.1 carrying a short FLAG-HA epitope (e-H3.1) under control of the *HTR9* endogenous promoter. Immunofluorescence staining of cotyledon nuclei showed e-H3.1 protein at the few visible conspicuous chromocenter structures at 2 dag and a clear enrichment at mature chromocenters at 5 dag (Fig. 3A).

To assess e-H3.1 enrichment at different genomic regions at 2 and 5 dag in a quantitative manner we carried out ChIP. At 5 dag, the e-H3.1 levels at 180 bp and TSI repeats are about 10-fold higher than at the active genes *HXK1* and *UEV1C* (Fig. 3C).

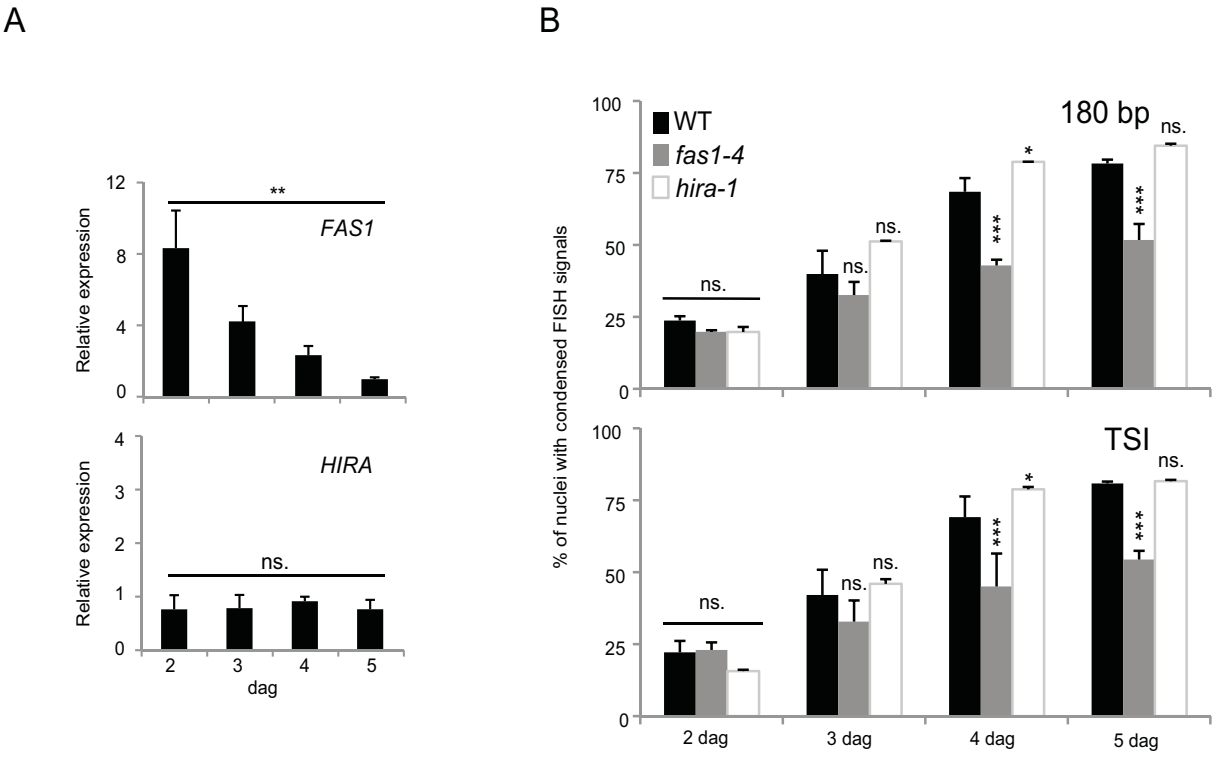


Figure 4: CAF-1 but not HIRA is required for chromocenter formation.

(A) RT-qPCR analysis of FAS1 and HIRA transcripts in WT cotyledons aged 2, 3, 4 or 5 dag. Histograms show mean fold change relative to transcript levels at 5 dag (set to 1) after normalization to At2g28390 expression ± SEM obtained for two independent PCR amplifications of two biological replicates. **p<0.01; Student's t-test. (B) Percentage of nuclei (± SEM from two independent biological replicates) showing clustering of 180 bp (top) and TSI (bottom) repeat sequences into chromocenter structures in fas1-4 and hira-1 mutant cotyledon nuclei compared to WT at 2, 3, 4 or 5 dag, n>200. *p<0.05; **p<0.01; ***p<0.005; $\chi 2$ test.

Interestingly, e-H3.1 enrichment is dynamic during this developmental time-window and increases at heterochromatic sequences between 2 and 5 dag, while the level remains stable at the active genes analyzed (Fig. 3C). This pattern was confirmed in an independent transgenic line expressing HTR1-CFP, another H3.1 encoding gene (Ingouff et al., 2010) (Fig. 3D). Therefore, H3.1 dynamics are distinct at heterochromatic and euchromatic loci during post-germination development and H3.1 becomes enriched at centromeric and pericentromeric repeats during chromocenter formation. This suggests a role for precise and site-specific deposition of the canonical H3.1 in the establishment of chromocenters as specialized chromatin domains.

Chromatin assembly mediated by CAF-1 is required for chromocenter formation

In mammals, the canonical H3.1 is deposited by the Chromatin Assembly Complex-1 (CAF-1) that operates in a DNA-synthesis dependent manner, while the Histone Regulator (HIR) complex deposits the replacement variant H3.3 (Tagami et al., 2004). The genes encoding FAS1, a subunit of the Arabidopsis CAF-1 complex, and HIRA are expressed during early post-germination development in cotyledons (Fig. 4A) and Arabidopsis mutants lacking functional CAF-1 and HIR complexes are viable (Duc et al., 2015; Kaya et al., 2001; Nie et al., 2014), allowing to test the implication of the two assembly complexes in chromocenter dynamics. FISH analysis for 180 bp and TSI repeats revealed that, while at 2 dag around 20% of nuclei show entirely clustered repeats in both WT and *fas1-4* mutants, clustering is observed for only 50% of the *fas1-4* mutant nuclei compared to ~80% in WT or *hira-1* at 5 dag (Fig. 4B). Defects in histone deposition mediated by HIRA therefore do not impact chromocenter formation, while it is affected in a CAF-1 complex mutant.

To know to what extent H3.1 enrichment and dynamics during post-germination development are altered in CAF-1 complex mutants, we introgressed the transgenic line expressing e-H3.1 into the *fas1-4* mutant and analyzed H3.1 occupancy dynamics (Fig. 5). Despite reduced transgene expression in the *fas1-4* mutant background, we observed significant nucleosomal e-H3.1 protein levels (Fig. 5A), which can be explained by the remaining CAF-1 activity in the *fas1-4* mutant allele (Duc et al., 2015; Ramirez-Parra and Gutierrez, 2007b). We found that, in contrast to WT plants in which e-H3.1 occupancy at 180 bp, TSI and the Ta3 retrotransposon increases between 2 and 5 dag, this dynamic is lost in *fas1-4* plants, illustrating that the CAF-1 histone chaperone complex contributes to H3.1 dynamics in particular at heterochromatic regions (Fig. 5A). Given the reported preferential enrichment of H3.1 in repressive marks, we also analyzed the dynamics of H3K27me1 and H3K9me2 enrichment in *fas1-4* mutants. The profile of H3K27me1 enrichment in *fas1-4* cotyledons at both euchromatic and heterochromatic targets is similar to WT (Fig. 2C and Fig. 5B). However, in contrast to WT, no increase in H3K9me2 was observed at heterochromatic repeats in *fas1-4* mutants between 2 and 5 dag (Fig. 2D and Fig. 5C), suggesting that loss of CAF-1 also disturbs H3K9me2 dynamics at 180 bp and TSI repeats.

Discussion

Cytological studies have provided significant insight into the different types of heterochromatic repeats that are organized into chromocenters and the dynamics of this higher order organization during

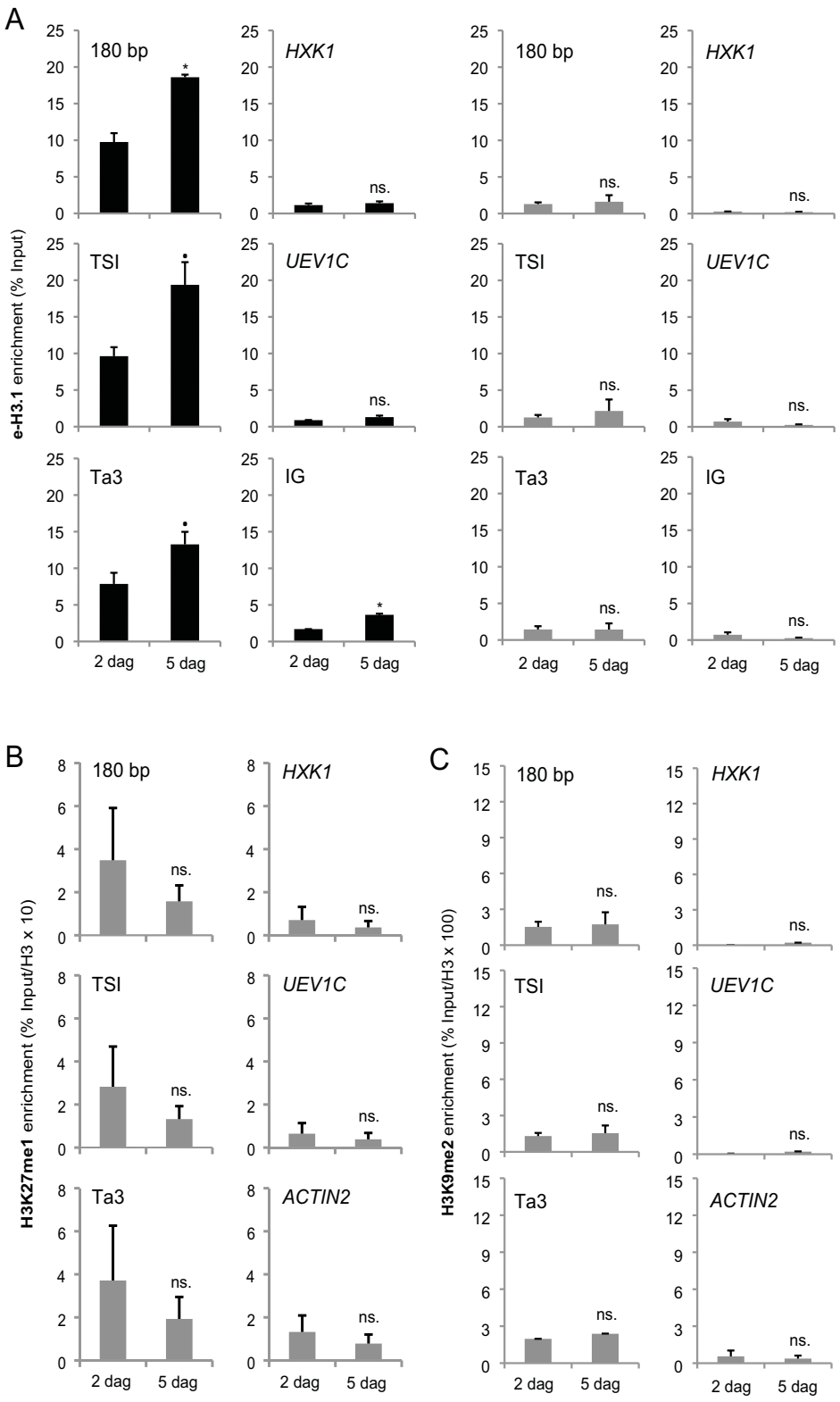


Figure 5: H3.1 and H3K9me2 dynamics at heterochromatin during chromocenter formation require the CAF-1 complex.

(A) Enrichment of e-H3.1 relative to input in cotyledons of e-H3.1/WT (black) or e-H3.1/fas1-4 plants (grey) at 180 bp, TSI and the Ta3 retrotransposon, two euchromatic loci (HXK1 and UEV1C) and an intergenic region at 2 and 5 dag as determined by ChIP-qPCR. Histograms present mean percentages ± SEM for at least 2 biological replicates. (B-C) Enrichment of H3K27me1 (C) or H3K9me2 (D) assessed by ChIP qPCR at 180 bp, TSI, the Ta3 retrotransposon, as well as HXK1, UEV1C and ACTIN2 at 2 or 5 dag in e-H3.1/fas1-4 cotyledons. Histograms show mean percentage ± SEM of the respective immunoprecipitation relative to H3 levels for three biological replicates. •p<0.1; *p<0.05; Student's t-test.

development. The mechanisms involved in chromocenter formation, however, and the local chromatin features of heterochromatic sequences at different states of higher order organization, have not been investigated. During post-germination development in cotyledons, decondensed centromeric and pericentromeric repeats are dispersed in the nucleus and then progressively cluster into chromocenter structures. Our analysis established that despite dispersion of these repeats at 2 dag, centromeric and pericentromeric repeats already display elevated nucleosomal occupancy and enrichment in repressive histone modifications compared to euchromatin. While we could not find significant changes in H3K27me1 methylation levels during chromocenter formation, the appropriate setting of this mark is required for chromocenter formation during cotyledon development, as clustering of centromeric and pericentromeric repeats is severely impaired in the H3K27 monomethyltransferase mutant *atxr5 atxr6* at late stages of chromocenter formation. In contrast, levels of another repressive chromatin mark, H3K9me2, further increase at heterochromatic regions during 2 and 5 dag. This may reflect global changes in H3K9me2 levels during this time window since euchromatic sequences are also affected and correlates with reduced TSI and Ta3 transcript levels at 5 dag compared 2 dag. Loss of the three H3K9 methyltransferases SUVH4, SUVH5 and SUVH6 caused an increased percentage of nuclei with dispersed centromeric and pericentromeric repeats at 2 dag, suggesting a potential role in maintenance of chromocenter organization during imbibition/germination. However, chromocenter formation then takes place normally in *suvh4 suvh5 suvh6* triple mutants and only clustering of the pericentromeric TSI repeats is affected at 5 dag. Even though histone methylation is strongly reduced in the *suvh4 suvh5 suvh6* triple mutants (Johnson et al., 2004; Yelagandula et al., 2014), a role for H3K9 methylation catalyzed by alternative histone methyltransferases in chromocenter formation cannot be excluded. Taken together, both repressive marks are enriched even at dispersed centromeric and pericentromeric repeats and differentially affect the clustering of these repeats in chromocenters.

Besides histone post-translational modifications, our data show that during chromocenter formation, the levels of canonical histone H3.1 increase specifically at heterochromatic regions. These dynamics depend on CAF-1, arguing for an important role for the histone chaperone complex CAF-1 in H3.1 dynamics at heterochromatic sequences and in the formation of higher-order chromatin structures. Indeed, the number of nuclei with fully formed chromocenter domains is reduced in *fas1-4* mutant nuclei echoing the requirement for CAF-1 in formation of higher-order chromatin domains during early mouse development (Akiyama et al., 2011; Houlard et al., 2006). Failure to assemble H3.1 at heterochromatin, however, might not be the only explanation for defective chromocenter formation upon loss of CAF-1, as replication-coupled deposition of histone modifications could be affected such as the timely H3K27 monomethylation of H3.1 that takes place in a replication-coupled manner (Jacob et al., 2014; Jacob and Feng, 2009). While we did not detect significant differences in H3K27me1 levels between WT and *fas1-4* mutants in cotyledon nuclei in this study, an observation that might be explained by residual CAF-1 activity in the *fas1-4* mutant (Ramirez-Parra and Gutierrez, 2007b) or

monomethylation by alternative and/or unknown H3K27 methyltransferases, we observed differences in H3K9me2 dynamics. Therefore, proper chromatin assembly at the replication fork by CAF-1 might contribute to maintenance and/or *de novo* setting of H3K9me2. We conclude that defective chromatin assembly in the absence of a functional CAF-1 complex impairs H3.1 deposition at heterochromatic repeats, the proper setting of epigenetic marks and ultimately chromocenter formation.

Cell proliferation in Arabidopsis is transiently activated after emergence of the cotyledons from the seed coat (Masubelele et al., 2005) and cells undergo endoreplication during cotyledon growth (Ramirez-Parra and Gutierrez, 2007b), allowing CAF-1-mediated H3.1 incorporation during this time window. To explain the increase in H3.1 specifically at heterochromatin, we can envisage that, in plants, H3.1 is deposited explicitly at heterochromatic repeats, possibly facilitated by specific timing of heterochromatin replication in S-phase. Alternatively, H3.1 is incorporated in a genome-wide manner, but further exchanged for H3.3 at the euchromatic loci analyzed here, leading to the observed differences between heterochromatic and euchromatic regions. Both scenarios suggest that at 2 dag, heterochromatin was depleted of H3.1 and enriched in other H3 variants.

Conspicuous chromocenter structures are found in several organisms and our data suggest that the formation of chromocenters reinforces transcriptional repression. Moreover, higher-order chromatin compaction may help maintaining other heterochromatin features, such as late replication timing, or contribute to minimizing recombination between sequences by restricting the possibility to interact (Quivy et al., 2004). Another intriguing hypothesis is that heterochromatin reorganization during germination and the subsequent seedling development is essentially driven by the need to reorganize euchromatic chromatin loops to ensure appropriate gene expression during the transition from heterotrophic to autotrophic growth. Dynamic reorganization of another pericentromeric repetitive array, the 5S rRNA genes, also occurs during this time window and was suggested to contribute to the sequestration and transcriptional repression of mutated 5S genes (Benoit et al., 2013; Douet et al., 2008; Mathieu et al., 2003a). While we concentrated here on histone H3 variants in chromocenter formation, the linker histone H1 that promotes higher-order chromatin structures (Robinson and Rhodes, 2006) or other core histone variants such as H2A.W that drives higher order chromatin organization (Yelagandula et al., 2014) might be important candidates besides H3.1 in chromocenter formation during this developmental stage.

Methods

Plant material and plasmid construction

Homozygous mutants (*fas1-4* (SAIL-662-D10), *hira-1* (WiscDsLox362H05), *htr9-1* (SALK_148171), *atxr5* (SALK_130607), *atxr6* (SAIL_181_D09) and *suvh4 suvh5 suvh6* (SALK_041474, Gabi_263C05, SAIL_1244_F04)) are in the Columbia background and were identified by PCR-based genotyping. After 2 days of stratification at 4ºC in the dark, seedlings were grown under 16 h light/8 h dark cycles at 23ºC on 1x MS medium. Cotyledons were harvested at 78 h, 102 h, 126 h or 150 h after transfer to the growth chamber, corresponding to time points 2, 3, 4 or 5 dag, respectively.

To generate the transcriptional fusion of H3.1 with the FLAG-HA tag, we cloned the OCS (octopine synthase) terminator, the FLAG-HA tag and the genomic fragment containing the promoter and the genomic coding region of *HTR9* (stop codon excluded) using Gateway technology. T3 monolocus homozygous lines were selected based on segregation of the resistance character.

Primers

Sequences of the primers used in this study are available upon request.

RNA extraction and RT-PCR

Total RNAs were extracted from dissected cotyledons using Tri-Reagent (Euromedex). Reverse transcription was primed with oligo(dT)15 or random hexamers supplemented with reverse primers for TSI and Ta3 using M-MLV reverse transcriptase (Promega). Transcript levels were determined by quantitative PCR with the LightCycler® 480 SYBR Green I Master kit on the Roche LightCycler® 480 and normalized to At2g28390 (Czechowski et al., 2005).

Fluorescence in situ hybridization and Immunofluorescence

More than 200 nuclei subjected to Fluorescence in situ hybridization (FISH) (Probst et al., 2003) were counted per condition using a double blind experimental setup. Only nuclei in which all 180 bp and TSI repeats are clustered in chromocenters are scored as “clustered”. Immunostainig with anti-HA (ab9110, Abcam) was performed as described (Pontes et al., 2006).

ChIP analysis

Chromatin was isolated as described (Bowler et al., 2004) and ChIP carried out with the LowCell# ChIP kit (Diagenode) followed by qPCR quantification relative to input. Anti-H3 (ab1791, Abcam), anti-H3K9me2 (ab1220, Abcam), anti-H3K4me3 (04-745, Millipore), anti-H3K27me1 (pAB-045-50, Diagenode), anti-GFP and FLAG-coupled magnetic beads (Sigma) were used.

Acknowledgments

We thank S. Jacobsen and C. Pikaard for seeds, S. Tourmente for stimulating discussion, A. Eccher, M. Peyny, G. Pompeu and E. Vanrobays for technical help and O. Mittelsten Scheid and J. Paszkowski for critical reading.

Author contributions

AVP and MB designed the research. MB, LS, CD, SC, MCE, SLG and AVP performed the experiments. MB, LS, CD, SC, SLG, AP and AVP analyzed the data. MB, CT and AVP wrote the manuscript.

Funding:

M.B., A.P. and L.S. are supported by stipends from the Region Auvergne and M.B. by an ARC fellowship. C.D., S.C., S.L.G. and A.V.P. were supported by ANR-11 JSV2 009 01 and ANR-12 ISV6 0001. CNRS, INSERM and Clermont Universities supported this work.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Akiyama T, Suzuki O, Matsuda J, Aoki F (2011) Dynamic replacement of histone H3 variants reprograms epigenetic marks in early mouse embryos. PLoS Genet **7**: e1002279

Almouzni G, Probst A V (2011) Heterochromatin maintenance and establishment: lessons from the mouse pericentromere. Nucleus **2**: 332–8

Benoit M, Layat E, Tourmente S, Probst A V (2013) Heterochromatin dynamics during developmental transitions in Arabidopsis - a focus on ribosomal DNA loci. Gene **526**: 39–45

Bowler C, Benvenuto G, Laflamme P, Molino D, Probst A V, Tariq M, Paszkowski J (2004) Chromatin techniques for plant cells. Plant J **39**: 776–789

Chodavarapu RK, Feng S, Bernatavichute Y V, Chen PY, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, et al (2010) Relationship between nucleosome positioning and DNA methylation. Nature **466**: 388–392

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol **139**: 5–17

Douet J, Blanchard B, Cuvillier C, Tourmente S (2008) Interplay of RNA Pol IV and ROS1 during post-embryonic 5S rDNA chromatin remodeling. Plant Cell Physiol **49**: 1783–1791

Duc C, Benoit M, Le Goff S, Simon L, Poulet A, Cotterell S, Tatout C, Probst A V. (2015) The histone chaperone complex HIR maintains nucleosome occupancy and counterbalances impaired histone deposition in CAF-1 complex mutants. Plant J. doi: 10.1128/AEM.01604-07

Ebbs M, Bender J (2006) Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. Plant Cell **18**: 1166–1176

Feng S, Cokus SJ, Schubert V, Zhai J, Pellegrini M, Jacobsen SE (2014) Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in Arabidopsis. Mol Cell 1–14

Fransz P, ten Hoopen R, Tessadori F (2006) Composition and formation of heterochromatin in Arabidopsis thaliana. Chromosome Res **14**: 71–82

Fransz P, de Jong JH, Lysak M, Castiglione MR, Schubert I (2002) Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate. Proc Natl Acad Sci U S A **99**: 14584–14589

Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, et al (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell **140**: 678–691

Grob S, Schmid MW, Grossniklaus U (2014) Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. Mol Cell 1–16

Houlard M, Berlivet S, Probst A V, Quivy J-PP, Héry P, Almouzni G, Gérard M (2006) CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. PLoS Genet **2**: e181

Ingouff M, Rademacher S, Holec S, Soljić L, Xin N, Readshaw A, Foo SH, Lahouze B, Sprunck S, Berger F (2010) Zygotic Resetting of the HISTONE 3 Variant Repertoire Participates in Epigenetic Reprogramming in Arabidopsis. Curr Biol **20**: 2137–2143

Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature **416**: 556–560

Jacob Y, Bergamin E, Donoghue MTA, Mongeon V, LeBlanc C, Voigt P, Underwood CJ, Brunzelle JS, Michaels SD, Reinberg D, et al (2014) Selective Methylation of Histone H3 Variant H3.1 Regulates Heterochromatin Replication. Science **343**: 1249–1253

Jacob Y, Feng S, LeBlanc CA, Bernatavichute Y V, Stroud H, Cokus S, Johnson LM, Pellegrini M, Jacobsen SE, Michaels SD (2009) ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat Struct Mol Biol **16**: 763–768

Jin C, Felsenfeld G (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes Dev **21**: 1519–1529

Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE (2004) Mass spectrometry analysis of Arabidopsis histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res **32**: 6511–6518

Kaya H, Shibahara KI, Taoka KI, Iwabuchi M, Stillman B, Araki T (2001) FASCIATA genes for chromatin assembly factor-1 in arabidopsis maintain the cellular organization of apical meristems. Cell **104**: 131–142

Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol **13**: 1192–1200

Loyola A, Bonaldi T, Roche D, Imhof A, Almouzni G (2006) PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. Mol Cell **24**: 309–316

Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J **24**: 800–812

Masubelele NH, Dewitte W, Menges M, Maughan S, Collins C, Huntley R, Nieuwland J, Scofield S, Murray JA (2005) D-type cyclins activate division in the root apex to promote seed germination in Arabidopsis. Proc Natl Acad Sci U S A **102**: 15694–15699

Mathieu O, Jasencakova Z, Vaillant I, Gendrel A V, Colot V, Schubert I, Tourmente S (2003) Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in Arabidopsis. Plant Cell **15**: 2929–2939

Mathieu O, Probst A V, Paszkowski J (2005) Distinct regulation of histone H3 methylation at lysines 27 and 9 by CpG methylation in Arabidopsis. EMBO J **24**: 2783–91

May BP, Lippman ZB, Fang Y, Spector DL, Martienssen RA (2005) Differential regulation of strand-specific transcripts from Arabidopsis centromeric satellite repeats. PLoS Genet **1**: e79

Mayer R, Brero A, von Hase J, Schroeder T, Cremer T, Dietzel S (2005) Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. BMC Cell Biol. 6:

Mito Y, Henikoff JG, Henikoff S (2005) Genome-scale profiling of histone H3.3 replacement patterns. Nat Genet **37**: 1090–1097

Nie X, Wang H, Li J, Holec S, Berger F (2014) The HIRA complex that deposits the histone H3.3 is conserved in Arabidopsis and facilitates transcriptional dynamics. Biol Open 1–9

Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Mittelsten Scheid O (2010) Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. Plant Cell **22**: 3118–3129

Pontes O, Li CF, Nunes PC, Haag J, Ream T, Vitins A, Jacobsen SE, Pikaard CS (2006) The Arabidopsis chromatin-modifying nuclear siRNA pathway involves a nucleolar RNA processing center. Cell **126**: 79–92

Probst A V., Fransz PF, Paszkowski J, Mittelsten Scheid O (2003) Two means of transcriptional reactivation within heterochromatin. Plant J **33**: 743–749

Quivy JP, Roche D, Kirschner D, Tagami H, Nakatani Y, Almouzni G (2004) A CAF-1 dependent pool of HP1 during heterochromatin duplication. EMBO J **23**: 3516–3526

Ramirez-Parra E, Gutierrez C (2007) E2F regulates FASCIATA1, a chromatin assembly gene whose loss switches on the endocycle and activates gene expression by changing the epigenetic status. Plant Physiol **144**: 105–20

Robinson PJJ, Rhodes D (2006) Structure of the “30 nm” chromatin fibre: a key role for the linker histone. Curr Opin Struct Biol **16**: 336–43

Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, et al (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. EMBO J **30**: 1928–38

Shu H, Wildhaber T, Siretskiy A, Gruissem W, Hennig L (2012) Distinct modes of DNA accessibility in plant chromatin. Nat Commun **3**: 1281

Steimer A, Amedeo P, Afsar K, Fransz P, Mittelsten Scheid O, Paszkowski J (2000) Endogenous targets of transcriptional gene silencing in Arabidopsis. Plant Cell **12**: 1165–1178

Stroud H, Otero S, Desvoyes B, Ramírez-Parra E, Jacobsen SE, Gutierrez C (2012) Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. Proc Natl Acad Sci **109**: 5370–5

Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y (2004) Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell **116**: 51–61

Tessadori F, Chupeau MC, Chupeau Y, Knip M, Germann S, van Driel R, Fransz P, Gaudin V (2007) Large-scale dissociation and sequential reassembly of pericentric heterochromatin in dedifferentiated Arabidopsis cells. J Cell Sci **120**: 1200–1208

Volle C, Dalal Y (2014) Histone variants: the tricksters of the chromatin world. Curr Opin Genet Dev **25C**: 8–14

Wollmann H, Holec S, Alden K, Clarke N, Jacques P, Berger F (2012) Dynamic deposition of histone variant h3.3 accompanies developmental remodeling of the Arabidopsis transcriptome. PLoS Genet **8**: e1002658

Yelagandula R, Stroud H, Holec S, Zhou K, Feng S, Zhong X, Muthurajan UM, Nie X, Kawashima T, Groth M, et al (2014) The Histone Variant H2A.W Defines Heterochromatin and Promotes Chromatin Condensation in Arabidopsis. Cell **158**: 98–109

Van Zanten M, Koini MA, Geyer R, Liu Y, Brambilla V, Bartels D, Koornneef M, Fransz P, Soppe WJ (2011) Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation. Proc Natl Acad Sci U S A **108**: 20219–24

VI. Identification et caractérisation de complexes impliqués dans la dynamique des histones

Bien que l'on connaisse de plus en plus de marques épigénétiques et de plus en plus de mécanismes de régulation du dépôt, éviction et lecture de ces marques, peu de données sont disponibles en ce qui concerne la dynamique des histones chez les plantes. Nous avons donc cherché à définir et caractériser des complexes protéiques impliqués dans ce processus.

A. Le complexe HIR maintient l'occupation nucléosomale et compense le défaut de dépôt des histones dans le mutant *caf-1*

Dans l'objectif de comprendre les mécanismes moléculaires impliqués dans la maintenance de la chromatine et dans la dynamique des histones, notre équipe a caractérisé le complexe HISTONE REGULATOR (HIR) dans l'organisme modèle *Arabidopsis thaliana*. Ce complexe est impliqué dans l'assemblage de la chromatine de manière réplication-indépendante et est conservé de la levure jusqu'à l'homme.

Chez *Arabidopsis*, seule la protéine HIRA appartenant à ce complexe a été identifiée (Ingouff et al., 2010), mais son rôle dans la dynamique des histones n'a pas été étudié.

L'équipe a identifié les 3 orthologues des autres sous unités du complexe HIR encore non décrit : NUCLEIN1, NUCLEIN2, CABIN1 et a validé et caractérisé les mutants pour chacune de ces sous unités et pour HIRA. Cette étude a permis de démontrer le rôle central de la protéine HIRA dans le complexe HIR car seuls les mutants *hira* présentent une fertilité réduite. Au niveau moléculaire, la perte de la protéine HIRA provoque une réduction de la densité nucléosomale aussi bien au niveau de l'euchromatine que de l'hétérochromatine. Au contraire, la perte du complexe CHROMATIN ASSEMBLY FACTOR-1 (CAF-1), responsable de l'assemblage des nucléosomes couplés à la réplication de l'ADN, entraine une diminution de la densité nucléosomale uniquement au niveau de l'hétérochromatine. Cette diminution de la densité nucléosomale au niveau de l'hétérochromatine chez les mutants *caf-1* et *hir* est associée à une levée du silencing de certains transposons.

Dans les plantes déficientes en complexes CAF-1 et HIR, on observe de lourds problèmes de développement couplés à une stérilité des plantes. Notre hypothèse de travail prédit que les complexes HIR et CAF-1 ont des rôles distincts mais partiellement complémentaires dans le dépôt des histones et que le complexe HIR permet de compenser la perte du complexe CAF-1 en restaurant partiellement la densité nucléosomale.

J'ai participé à l'élaboration de ce manuscrit en déterminant les niveaux de ploïdie des noyaux de feuilles de différents mutants par des expériences de cytométrie en flux ainsi que par l'évaluation de la fertilité des plantes mutantes et des plantes issues de croisements entre les différents mutants pour les sous-unités du complexe CAF-1 et HIR par comptage de graines viables et avortées dans les siliques.

the plant journal

SEB Society for Experimental Biology

The Plant Journal (2015) **81,** 707–722 doi: 10.1111/tpj.12758

The histone chaperone complex HIR maintains nucleosome occupancy and counterbalances impaired histone deposition in CAF-1 complex mutants

Céline Duc, Matthias Benoit, Samuel Le Goff, Lauriane Simon, Axel Poulet, Sylviane Cotterell, Christophe Tatout and Aline V. Probst*

Génétique, Reproduction et Développement, CNRS UMR 6293, Clermont Université, INSERM U1103, 24 Avenue des Landais, BP 80026, Aubière Cedex 63171, France

Received 3 September 2014; revised 21 November 2014; accepted 23 December 2014; published online 19 January 2015.
*For correspondence (e-mail aline.probst@univ-bpclermont.fr).

SUMMARY

Chromatin organization is essential for coordinated gene expression, genome stability, and inheritance of epigenetic information. The main components involved in chromatin assembly are specific complexes such as Chromatin Assembly Factor 1 (CAF-1) and Histone Regulator (HIR), which deposit histones in a DNA synthesis-dependent or -independent manner, respectively. Here, we characterize the role of the plant orthologs Histone Regulator A (HIRA), Ubinuclein (UBN) and Calcineurin Binding protein 1 (CABIN1), which constitute the HIR complex. Arabidopsis loss-of-function mutants for the various subunits of the complex are viable, but *hira* mutants show reduced fertility. We show that loss of HIRA reduces extractable histone H3 protein levels and decreases nucleosome occupancy at both actively transcribed genes and heterochromatic regions. Concomitantly, HIRA contributes to maintenance of silencing of pericentromeric repeats and certain transposons. A genetic analysis based on crosses between mutants deficient in subunits of the CAF-1 and HIR complexes showed that simultaneous loss of both the CAF-1 and HIR histone H3 chaperone complexes severely affects plant survival, growth and reproductive development. Our results suggest that HIRA partially rescues impaired histone deposition in *fas* mutants to preserve nucleosome occupancy, implying plasticity in histone variant interaction and deposition.

Keywords: Arabidopsis, chaperone, histone, nucleosome, transcriptional silencing.

INTRODUCTION

Eukaryotic DNA is organized into chromatin. Its basic subunit, the nucleosome, consists of 146 bp of DNA wrapped around a histone octamer comprising a $(H3–H4)_2$ tetramer and two histone H2A–H2B dimers. Chromatin organization profoundly affects the accessibility of DNA to the cellular machinery, and therefore affects all cellular processes operating on DNA. To coordinate these different functions, remodeling of chromatin is required to allow access or exclusion of various factors. Remodeling can encompass movement of nucleosomes along the DNA, but also their disassembly and reassembly, as occurs during passage of the transcriptional machinery (Petesch and Lis, 2012). These processes are facilitated by factors modulating the stability of the nucleosomes or the association of histones with the DNA, such as covalent modifications and incorporation of histone variants (Jin and Felsenfeld, 2007). Except for histone H4, all histone proteins occur in non-canonical variants that differ in their primary amino acid sequence from the canonical paralogs (Talbert and Henikoff, 2010). These differences range from a few amino acids to large protein domains (Talbert *et al.*, 2012). For example, the canonical histone H3.1 and its variant H3.3 diverge by only four amino acids, but are incorporated differently during the cell cycle and show specific distribution patterns in mammals and plants (Tagami *et al.*, 2004; Goldberg *et al.*, 2010; Stroud *et al.*, 2012; Wollmann *et al.*, 2012; Filipescu *et al.*, 2013; Shu *et al.*, 2014). While nucleosomes containing H3.1 are thought to package DNA globally in a DNA synthesis-linked process, the replacement variant H3.3 is preferentially incorporated at enhancers, promoters and gene bodies of actively transcribed genes throughout the cell cycle (Ahmad and Henikoff, 2002; Jin *et al.*, 2009; Goldberg *et al.*, 2010).

The highly basic histone proteins are accompanied from synthesis to chromatin assembly by a network of histone chaperones (De Koning *et al.*, 2007), thereby preventing uncontrolled interaction with nucleic acids or negatively charged proteins. Consequently, histone chaperones are involved in all aspects of histone dynamics, such as transport and storage, chromatin assembly and disassembly, and parental histone transfer during DNA replication (Filipescu *et al.*, 2013; Groth *et al.*, 2007; De Koning *et al.*, 2007). Histone chaperones may be classified on the basis of their preferential binding to either H3–H4 or H2A–H2B subunits. In addition, some chaperones show specificity for particular histone variants and play a crucial role in their chromatin distribution (Tagami *et al.*, 2004; Drané *et al.*, 2010; Goldberg *et al.*, 2010).

While H3 histone variants are assumed to have evolved independently in animals and plants (Ingouff and Berger, 2010), histone H3 chaperones are highly conserved through evolution. The chaperone Anti-silencing function 1 (Asf1) binds H3–H4 dimers (English *et al.*, 2006; Natsume *et al.*, 2007) in the cytoplasm, and is involved in histone import into the nucleus (Campos *et al.*, 2010). Asf1 then transfers histones to chaperone complexes involved in nucleosome assembly. In mammals, two distinct pathways control deposition of either the canonical histone H3.1 or the variant H3.3. Chromatin Assembly Factor 1 (CAF-1), consisting of the three subunits p150, p60 and p48, ensures histone deposition in a DNA synthesis-dependent manner during replication and repair (Stillman, 1989; Gaillard *et al.*, 1996). CAF-1 specifically deposits H3.1 (Tagami *et al.*, 2004; Drané *et al.*, 2010) and interacts with Asf1 (Tyler and Collins, 2001). Independently of DNA synthesis and throughout the whole cell cycle, histone deposition is promoted by Histone Regulator A (HIRA), which shows high specificity for the variant H3.3 (Ray-Gallet *et al.*, 2002; Tagami *et al.*, 2004). HIRA depletion results in reduced genome-wide loading of H3.3 (Goldberg *et al.*, 2010; Pchelintsev *et al.*, 2013). In addition, H3.3 is deposited in mammals by death-associated protein (DAXX), α-thalassemia/mental retardation X-linked syndrome protein (ATRX) (Drané *et al.*, 2010; Goldberg *et al.*, 2010) and the chaperone DEK (Sawatsubashi *et al.*, 2010). HIRA is part of a multimeric complex termed the HIR complex, which was first identified in *Saccharomyces cerevisiae* in which it consists of four subunits (Hir1, Hir2, Hir3 and Hpc2) and functions as repressor of histone genes outside S phase (Osley and Lycan, 1987). Orthologs of Hir1 and Hir2 have been identified as HIRA in *Drosophila*, mammals and Arabidopsis (Phelps-Durr *et al.*, 2005; Nie *et al.*, 2014). In humans, the complex further comprises Ubinucleins 1 and 2 (UBN1 and UBN2), orthologs of yeast Hpc2, as well as the Calcineurin Binding protein (CABIN1), an ortholog of Hir3. HIRA mediates binding to UBN1 (Balaji *et al.*, 2009) and CABIN1 (Yang *et al.*, 2011) in mammals, and binds UBN2 in plants (Nie *et al.*, 2014). HIRA further interacts with Asf1 (Tang *et al.*, 2006; Nie *et al.*, 2014).

The study of histone chaperone complexes as well as their role in histone dynamics during development is hampered in higher organisms, as these factors are essential for survival. Mice embryos deficient in p150, the large subunit of the CAF-1 complex, fail to develop beyond early embryonic stages (Houlard *et al.*, 2006). Similarly, mutants of the *Drosophila* ortholog p180 die during larval development (Klapholz *et al.*, 2009). In Arabidopsis, the CAF-1 complex consists of the subunits FASCIATA1 (FAS1), FASCIATA2 (FAS2) and MULTICOPY SUPPRESSOR OF IRA1 (MSI1) (Kaya *et al.*, 2001). Arabidopsis mutants deficient in FAS1 or FAS2 are viable, but show pleiotropic morphological abnormalities, such as fasciated stems, serrated leaves and meristem alterations (Kaya *et al.*, 2001; Exner *et al.*, 2006; Kirik *et al.*, 2006). Consistent with its role in chromatin assembly during DNA replication, CAF-1 mutants fail to maintain repressive chromatin states, as illustrated by weak transcriptional reactivation of silent endogenous repetitive sequences and the stochastic reactivation of certain transposable elements (Takeda *et al.*, 2004; Ono *et al.*, 2006; Schönrock *et al.*, 2006).

Depletion of mammalian HIRA is also lethal, and HIRA knockout mice die during embryonic development (Roberts *et al.*, 2002). Furthermore, down-regulation of HIRA in *Xenopus* embryos causes gastrulation defects (Szenker *et al.*, 2012), phenotypes that may be explained by a role for HIRA in transcription (Formosa *et al.*, 2002; Schwartz and Ahmad, 2005; Ray-Gallet *et al.*, 2011). In contrast, *Drosophila* HIRA is only required for H3.3 deposition in the male pronucleus after fertilization (Loppin *et al.*, 2005), but not for viability (Bonnefoy *et al.*, 2007). How defective histone assembly outside S phase mediated by HIRA affects plants at the molecular level has not yet been addressed.

We show here that HIRA is an important player with respect to histone dynamics in Arabidopsis. Plants lacking HIRA are viable but show developmental defects and are impaired in the maintenance of transcriptional silencing. Loss of HIRA results in reduced H3 protein levels and affects nucleosome occupancy, not only at euchromatic but also at heterochromatic targets. Simultaneous loss of CAF-1 and HIR complexes severely affects Arabidopsis development. Surviving plants show significant defects in plant growth and reproduction, as well as an important reduction in nucleosome occupancy, without being further affected in maintenance of silencing and heterochromatin organization. Our data suggest that, in plants, the two evolutionarily conserved chromatin assembly complexes CAF-1 and HIR are involved in independent pathways of nucleosomal assembly, but show partial functional redundancy in maintenance of nucleosome occupancy.

RESULTS

Characterization of the Arabidopsis orthologs of HIR subunits

Arabidopsis orthologs of the HIR subunits are encoded by the following genes: *HIRA* (At3g44530), *UBN1* (At1g21610), *UBN2* (At1g77310) and *CABIN1* (At4g32820) (Nie *et al.*, 2014). We analyzed their protein sequences and observed conservation of the overall protein structure in Arabidopsis HIRA, which contains N-terminal WD40 repeats involved in protein–protein interaction, the C-terminal Hira domain and the B-domain required in mammals for binding to ASF1 (Figure 1a and Figure S1a,b) (Tang *et al.*, 2006). As in mammals, several plant species contain two closely related UBN paralogs (Figure 1a and Figure S1e). The Arabidopsis UBN proteins are of similar length, show 59.7% sequence identity, and conservation of the Hpc2-related domain (HRD) (Banumathy *et al.*, 2009) and the NHRD (N-terminal to the HRD region domain) (Figure S1c,d) (Tang *et al.*, 2012). The Arabidopsis HIR complex also comprises a third subunit, CABIN1, characterized by a range of tetratricopeptide (TPR)-like bi-helical repeats, which may form a scaffold for protein–protein interaction (Balaji *et al.*, 2009) (Figure 1a). Phylogenetic analysis confirmed that CABIN1 is found throughout the plant kingdom (Figure S1f). The conservation of specific domains and interaction of HIRA with Arabidopsis UBN2 and ASF1 (Nie *et al.*, 2014) suggests that, as in animals, HIRA is the scaffold protein for CABIN1 and UBN recruitment and binds ASF1.

To obtain insight into the biological function of the Arabidopsis HIR complex, we confirmed expression of all four genes in several tissues (Figure 1b and Figure S1g). We obtained T-DNA insertion mutants for each gene encoding a HIR complex subunit (Figure 1c–f). As conflicting results were reported concerning the viability of *hira* mutants (Phelps-Durr *et al.*, 2005; Ingouff *et al.*, 2010; Nie *et al.*, 2014), we obtained four T-DNA insertion lines and determined the exact insertion sites in the *HIRA* locus (Figure 1c and Figure S2a). We identified plants homozygous for all mutant alleles except SALK_143806, for which we did not identify any T-DNA insertion in various seed batches, but only the *hira-1* mutant (Ingouff *et al.*, 2010; Nie *et al.*, 2014) showed an effect on *HIRA* expression (Figure 1c and Figure S2b). We also generated a transgenic line expressing an artificial microRNA targeting *HIRA* transcripts ($hira^{amiRNA}$), and confirmed reduction of *HIRA* transcript levels to approximately 25% (Figure S2c). In addition to *hira* mutants, we identified T-DNA insertion alleles for *UBN1*, *UBN2* and *CABIN1*, and confirmed the absence of the corresponding full-length transcripts (Figure 1d–f). Single mutants of the HIR complex subunits display no obvious vegetative developmental phenotypes compared to wild-type (WT) plants (Figure 1g) except few cases of *hira-1* mutants with serrated leaves (Nie *et al.*, 2014). In comparison, *fas1-4* and *fas2-5* mutants, which possess mutations in one of the two larger subunits of the CAF-1 complex, show pleiotropic phenotypes (Figure S2d,e). Deficiency in the CAF-1 histone chaperone affects endoreplication levels, leading to a premature switch to the endocycle and increased endopolyploidy levels (Ramirez-Parra and Gutierrez, 2007). To evaluate endopolyploidy levels in HIR complex mutants, we performed flow-cytometry analysis. Whereas *fas2-5* mutant plants show a significantly increased proportion of 8C and 16C nuclei (one and two rounds of endoreplication), endopolyploidy profiles of HIR mutants are not significantly different from those of WT plants (Figure 1h). This suggests that absence of a functional HIR complex does not affect endoreplication.

As mutants impaired in histone deposition mediated by the CAF-1 complex show reduced fertility (Figure S2e,f) (Ramirez-Parra and Gutierrez, 2007), we looked more closely at the siliques of HIR complex mutants. Seed set in self-pollinated *hira-1* mutant flowers, but not in the other HIR complex mutants, was significantly reduced, and we observed more unfertilized ovules and aborted seeds than in WT (Figure 2a,b). A similar phenotype was noted in the $hira^{amiRNA}$ line (Figure 2c). We observed standard Mendelian genetic transmission of the T-DNA mutant allele in the progeny of self-fertilized *hira-1*/*HIRA* plants (51 heterozygous and homozygous *hira-1* plants, $n = 72$, $P < 0.05$), similar to previous reports (Ingouff *et al.*, 2010; Nie *et al.*, 2014). Nevertheless, while anthers of HIR complex mutants develop normally (Figure 2d), in contrast to the heart-shaped *fas1-4* and *fas2-5* anthers (Figure S2g), a proportion of pollen grains are non-viable in *hira-1*, as revealed by Alexander staining (Figure 2d), but to a lesser extent than in *fas* mutants (Figure S2g). Furthermore, an increased number of unfertilized ovules were observed in heterozygous *hira-1*/*HIRA* plants compared to their WT sister plants (Figure 2e). Taken together, these results suggest that, while the HIR complex is dispensable for vegetative growth, loss of the HIRA subunit causes sporophytic and gametophytic defects, resulting in reduced fertility.

The *hira-1* mutant shows a reduced histone H3 protein pool and altered nucleosome occupancy

In analogy to yeast and animal models, Arabidopsis CAF-1 and HIR complexes are assumed to bind non-nucleosomal histones and coordinate their assembly into nucleosomes. We therefore investigated whether mutations in these two complexes affect the pool of H3 histones. We extracted proteins as described by Durut *et al.* (2014), and determined the amount of histone H3 in *fas1-4* and *hira-1* mutants relative to WT plants by Western blotting (Figure 3a). We observed reduced H3 levels in both mutants, suggesting that histone flow and the amount available for

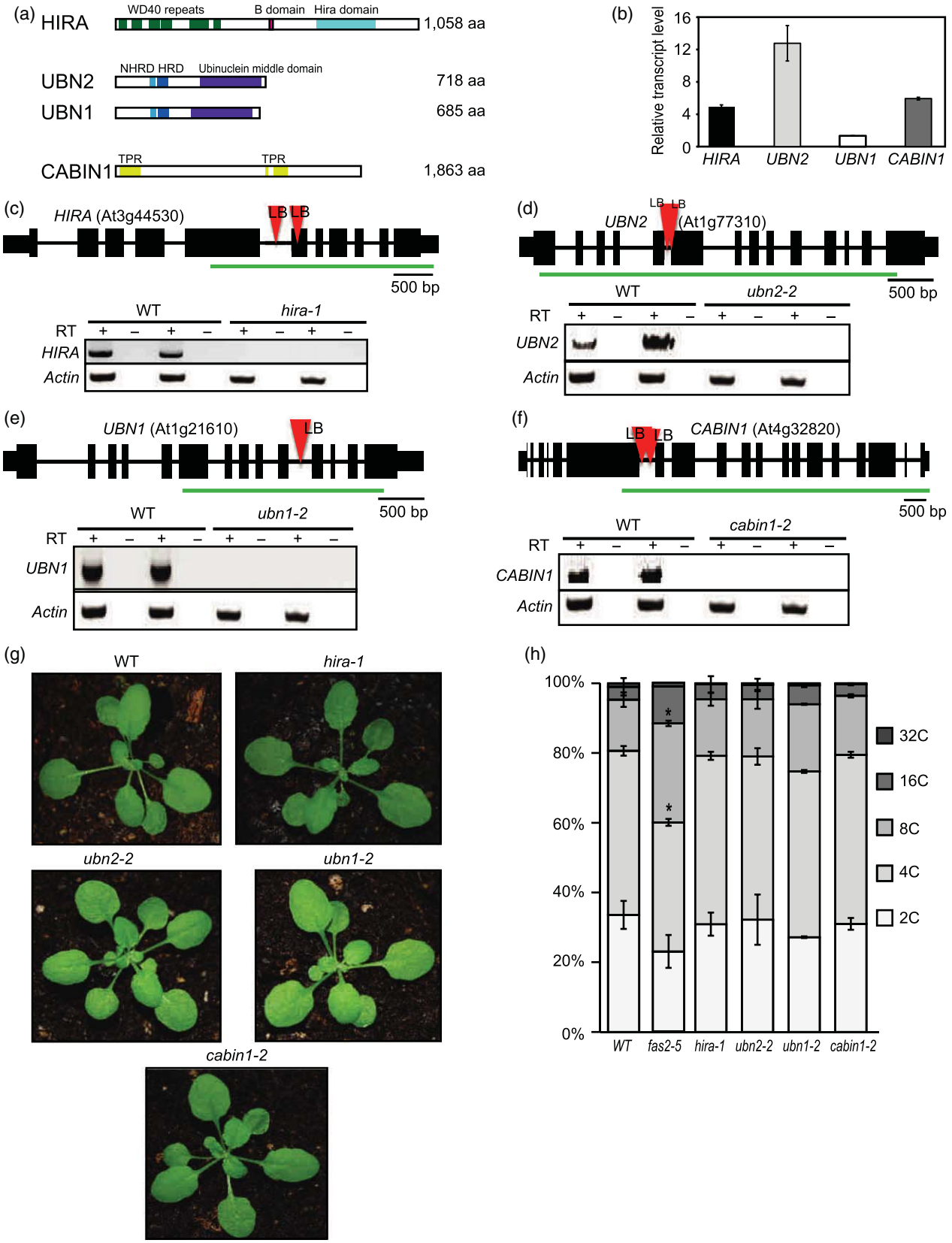


Figure 1. Characterization of the HIR complex subunits in plants.
(a) Functional domains of the HIR complex proteins. HRD, Hpc2-related domain; NHRD, N-terminal to the HRD region domain; TPR, tetratricopeptide-like bi-helical repeats.
(b) Quantitative RT-PCR analysis of *HIRA*, *UBN2*, *UBN1* and *CABIN1* transcript levels in 4-week-old plants.
(c–f) Gene structures and characterization of the Arabidopsis HIR complex mutants. Exons, boxes; UTRs, narrower boxes; introns, lines; T-DNA insertion, red triangle; LB, left border. Absence of full-length transcripts is revealed by RT-PCR on two biological replicates: the amplified region is indicated by a green line.
(g) Representative 25-day-old plantlets of HIR complex mutants grown on soil.
(h) Ploidy level distribution of CAF-1 and HIR complex mutant nuclei from 10-day-old plants.

de novo deposition are altered upon loss of CAF-1 or HIR chromatin assembly complexes.

To analyze nucleosome occupancy at specific genomic sites in WT, *fas1-4* and *hira-1* mutants in a quantitative manner, we used H3-ChIP combined with quantitative PCR. Based on one of the available whole-genome ChIP-Seq datasets (Stroud *et al.*, 2012), we selected euchromatic, intergenic and heterochromatic sites differentially enriched in H3.3 and H3.1 (Figure S3b–h). We first analyzed nucleosome occupancy at three constitutively active genes (*UBC28*, *UEV1C* and *HXK1*) with different expression levels (Figure S3a) that are enriched in H3.3 in their middle and 3′ regions (Figure S3b–d) (Stroud *et al.*, 2012), suggesting that they are preferential targets for H3 deposition mediated by HIRA, which co-immunoprecipitates with H3.3 (Nie *et al.*, 2014). In 3-week-old *fas1-4* mutant plants grown *in vitro*, nucleosome occupancy was unaffected, but in *hira-1* mutants, *UEV1C* and *HXK1* show mildly reduced H3 levels in the middle and 3′ regions (Figure 3b–d), but not in the 5′ region at which neither H3.1 nor H3.3 are enriched (Figure S3b–d). This difference is not due to altered nucleosome displacement associated with increased expression (Figure S3i), revealing a specific role for HIRA in H3 deposition at these genic regions.

As *fas1* mutants were previously shown to have moderately reduced nucleosome occupancy at selected pericentromeric sequences (Pecinka *et al.*, 2010), we also included heterochromatic repetitive elements (180 bp repeat, 106B centromeric satellites and an endogenous family of transcriptionally repressed repeats called transcriptionally silent information (TSI) (Steimer *et al.*, 2000), as well as an intergenic region (Pecinka *et al.*, 2010) in our analysis. These heterochromatic regions are enriched in H3.1 (Figure S3e–g), but neither H3.1 nor H3.3 are particularly enriched at the intergenic region (Figure S3h). We observed reduced nucleosomal occupancy at 106B and TSI in *fas1-4* mutants (Figure 3e). Unexpectedly, we also found that nucleosome occupancy is reduced in *hira-1* mutants in these two heterochromatic regions and in the intergenic region (Figure 3e).

We conclude that loss of function of either of the two chaperone complexes CAF-1 and HIR affects the extractable H3 histone pool. Furthermore, loss of HIRA affects nucleosome occupancy at both euchromatic and heterochromatic regions, while loss of CAF-1 mainly affects heterochromatic sequences.

Loss of HIRA interferes with maintenance of transcriptional silencing but not with gene induction upon salt stress

The observed differences in nucleosome occupancy prompted us to investigate the functional consequences of altered histone dynamics in HIR complex mutants. We first analyzed the impact on maintenance of transcriptional silencing at the heterochromatic regions by ChIP combined with quantitative PCR. Quantification of transcript levels by quantitative RT-PCR revealed partial silencing release of TSI in *fas* mutants, as expected. In agreement with the changes in nucleosome occupancy, the *hira-1* mutants also show alleviation of TSI silencing (Figure 4a), which is not seen in *ubn2-2*, *ubn1-2* and *cabin1-2* mutants. None of the mutants of HIR complex subunits reactivates silencing at 106B, 180 bp or a multicopy transgenic locus (Morel *et al.*, 2000) (Figure 4a and Figure S4a). To examine whether the silencing release in *hira-1* is restricted to TSI sequences or is more general, we tested additional targets and observed alleviation of silencing of the Ta3 retrotransposon and a Mutator-like DNA transposon (*Mule*, At2g15810) (Figure 4b) in *hira-1* compared to WT plants, concomitant with reduced nucleosome occupancy in these two regions (Figure 4c).

Given the suggested function of HIRA in transcription in other species, we wished to determine whether HIRA is required to rapidly activate gene expression in response to an environmental stimulus. We exposed plants to salt stress and analyzed the expression of four genes induced under this stress condition (Zeller *et al.*, 2009). Under normal growth conditions, *Protein Phosphatase 2C* (*PP2C*, At3g16800) and *Ethylene Responsive Factor/APETALA 2* (*ERF/AP2*, At1g74930) are moderately expressed, while *MYB domain protein 41* (*MYB41*, At4g28110) and *C-Repeat/DRE binding factor 1* (*CBF1*, At4g25490) are not expressed (Zeller *et al.*, 2009; Duc *et al.*, 2013). In WT plants, all four genes are induced after 1 h of exposure to high-salt medium, undergoing changes in transcript levels from twofold to several hundred times, depending on the gene (Figure 4d). However, the *hira-1* mutant plantlets are not impaired in the rapid transcriptional response, and upregulate salt-responsive genes similarly to WT (Figure 4d). We next analyzed three genes that are repressed by salt stress (Zeller *et al.*, 2009): At4g12510 and At4g12520, which are expressed at a low level under normal growth

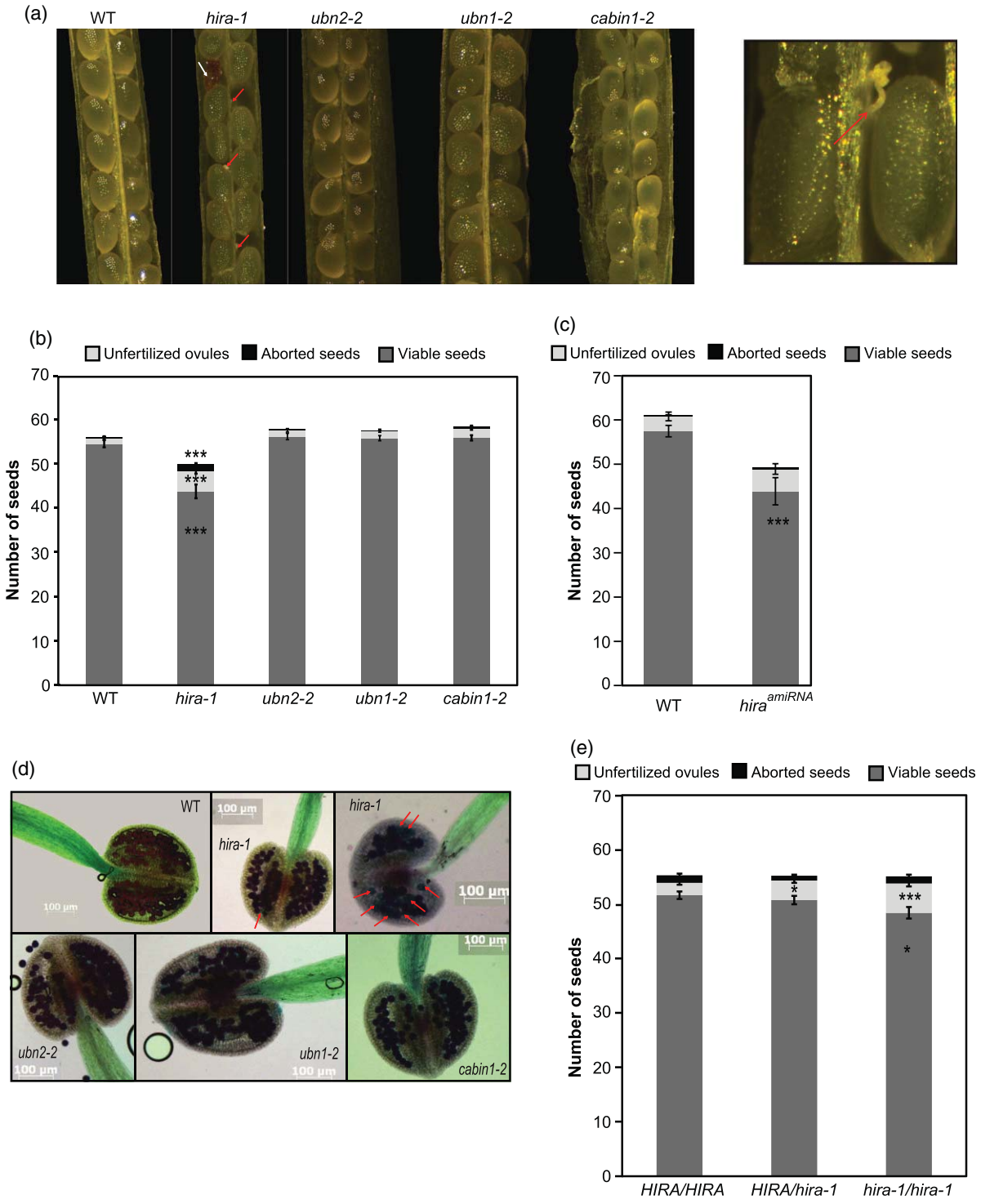
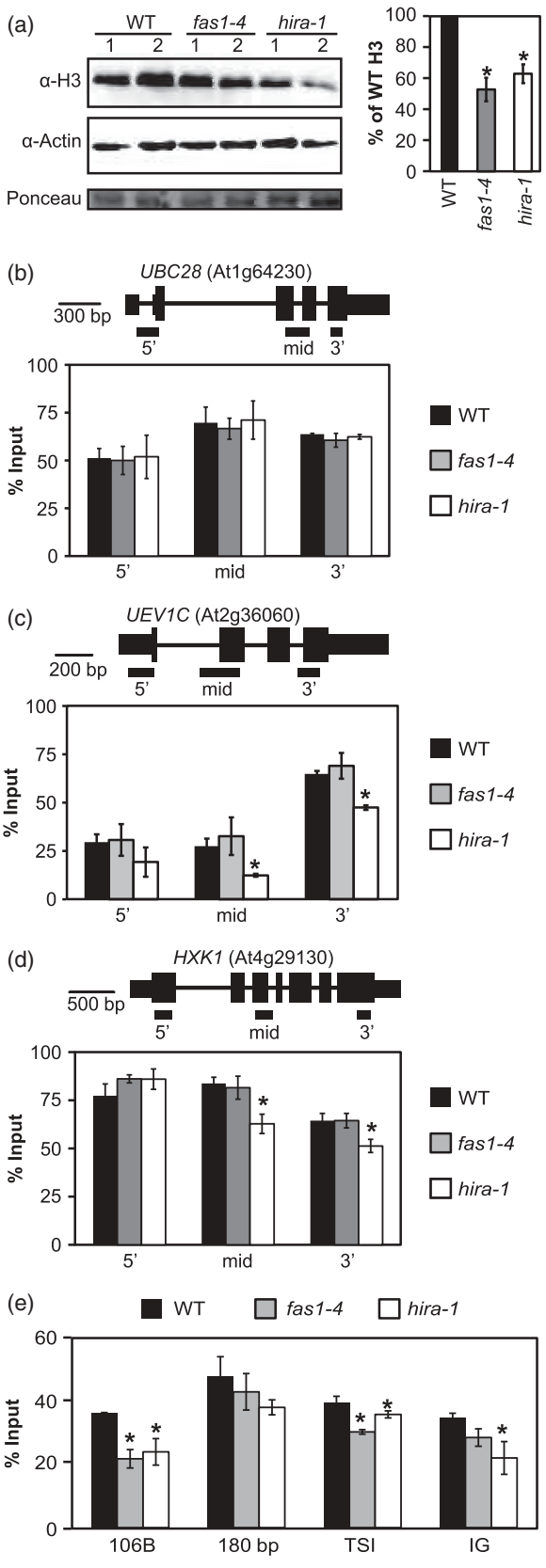


Figure 2. Analysis of Arabidopsis mutants in the HIR complex.
(a) Representative dissected siliques from HIR complex mutants. Red arrows indicate unfertilized ovules, and the white arrow indicates an aborted seed. Right, close-up of an unfertilized ovule.
(b) Quantification of seed content in HIR complex mutant siliques. Quantifications were obtained from 30 pooled siliques from at least four plants.
(c) Quantification of seed content in the *hira*^{amiRNA} line. Quantifications were obtained from 16 WT and 19 *hira*^{amiRNA} pooled siliques from four plants.
(d) Pollen viability assessed by Alexander staining. Only *hira-1* mutant anthers contain non-viable pollen (green color), as indicated by red arrows.
(e) Quantification of seed content in WT (*HIRA*/*HIRA*), heterozygous (*HIRA*/*hira-1*) and homozygous (*hira-1*/*hira-1*) siliques. Quantifications were obtained from 30 pooled siliques from at least six plants.



conditions, and *Small Auxin Upregulated RNA 14* (*SAUR14, At4g38840*), which is moderately expressed under normal growth conditions (Zeller *et al.*, 2009; Duc *et al.*, 2013). Similarly to salt-induced genes, *hira-1* mutant plantlets down-regulate the expression of these genes, as in WT (Figure 4e).

Taken together, HIRA is implicated in the maintenance of transcriptional gene silencing at a selection of endogenous repeat elements and transposons, but is dispensable for rapid gene expression changes of the examined genes upon salt exposure.

Genetic analyses identify HIRA as the central subunit of the HIR complex

To obtain further insight into the relative importance of the various HIR complex members and to examine the epistatic relationship between CAF-1 and HIR complexes, we first crossed *fas1* and *fas2* mutants with each mutant for the various HIR complex subunits (Figure S5a). We performed segregation analyses on F_2 progeny from several independent F_1 plants obtained for each cross by selecting homozygous *fas* plants which display serrated leaf margins and genotyping for the corresponding HIR complex mutation (Table 1). Double mutants of *cabin1-2*, *ubn2-2* or *ubn1-2* with *fas1-4* show the expected segregation ratio (Figure S5a and Table 1) worsening the *fas1-4* growth phenotypes and fertility defects. (Figure 5a,b and Figure S5b). In contrast, double mutants of *cabin1-2*, *ubn2-2* and *ubn1-2* with *fas2-5* were recovered with reduced frequency (Table 1) but were indistinguishable from single *fas2-5* mutants (Figure 5c and Figure S5c). Strong reduction in seed set was observed for *fas2-5 ubn2-2* and *fas2-5 cabin1-2* plants (Figure 5d and Figure S5d), and very little viable pollen was detected in anthers, which were frequently aberrantly shaped in *fas2-5 ubn2-2* plants (Figure S5e,f). In comparison with the other HIR complex subunits, the double mutants generated between *hira-1* and CAF-1 mutants revealed more severe phenotypes. Only a few *fas2-5 hira-1* plants were obtained with and without prior selection of the *fas2* phenotype (Table 1 and Table S1a), and their development was arrested before formation of the first leaves (Figure 5c). Furthermore, *fas2-5 hira-1*/*HIRA* plants

Figure 3. HIRA loss affects the extractable histone pool and nucleosome occupancy.
(a) Left: Histone H3 protein levels quantified by Western blotting. Twenty micrograms of proteins extracted from two independent biological replicates of 3-week-old *in vitro* grown plants were loaded per lane. Right: Quantification of H3 band intensities relatively to actin from three independent experiments.
(b–e) Histone H3 occupancy at various positions along three active genes [*UBC28* (b), *UEV1C* (c) and *HXK1* (d)] and at heterochromatic repeats and an intergenic region (e) assessed by H3-ChIP combined with quantitative PCR in 3-week-old *in vitro* grown WT, *fas1-4* and *hira-1* mutant plants.

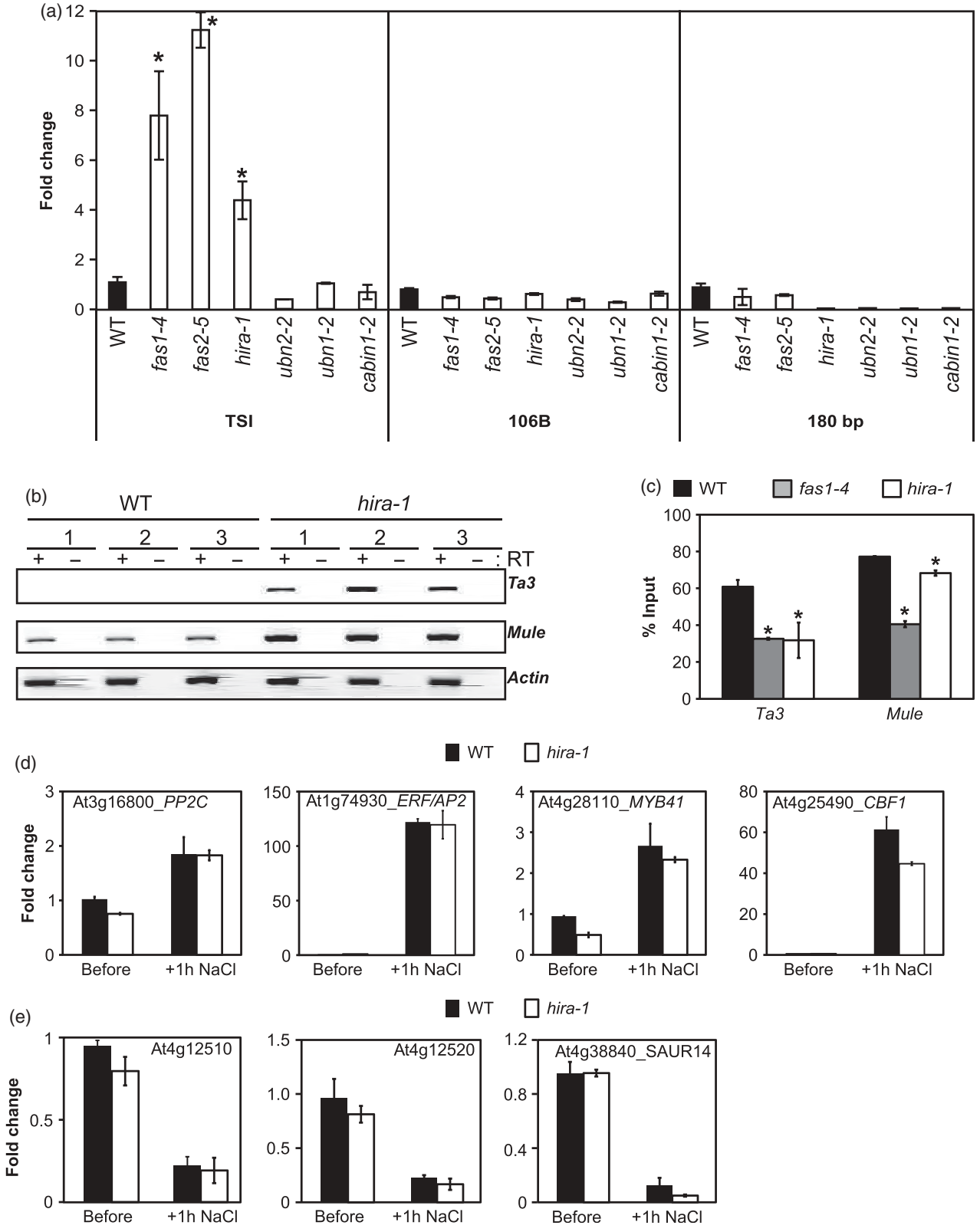


Figure 4. Release of transcriptional gene silencing and induction of gene expression in *hira-1* mutants.
(a) Quantitative RT-PCR analysis of TSI, 106B and 180 bp transcripts in 18-day-old CAF-1 and HIR complex mutant plants grown on soil.
(b) RT-PCR analysis of *Ta3* and *Mule* (At2g15810) transcripts of three independent biological replicates from WT and *hira-1* plants. *Actin* was used for normalization.
(c) Histone H3 occupancy at a heterochromatic pericentromeric region (*Ta3*) and an intergenic transposon (*Mule*, At2g15810) assessed by H3-ChIP combined with quantitative PCR in 3-week-old *in vitro*-grown WT, *fas1-4* and *hira-1* mutant plants.
(d,e) Quantitative RT-PCR analysis of transcript levels of salt-induced genes (*PP2C*, *ERF/AP2*, *MYB41* and CBF1) (d) and salt-repressed genes (At4g12510, At4g12520 and *SAUR14*) (e) before and after exposure of 10-day-old WT and *hira-1* mutant plantlets to 200 mM NaCl for 1 h (+1 h NaCl). The value for WT before salt treatment was set to 1.

were almost sterile (Figure S5g) and showed little viable pollen (Figure S5h). Similarly, significantly fewer *fas1-4 hira-1* double homozygous mutants were obtained (Table 1 and Table S1b). They develop beyond the state of *fas2-5 hira-1* mutant plants but are dwarf and dark green (Figure 5e), show flowers harboring mis-shapen carpels and short stamens with aberrant anthers (Figure 5f), and do not produce siliques. This finding is in agreement with the complete male sterility revealed by Alexander staining (Figure 5g), in comparison to their heterozygous sister plants (Figure S5i). We anticipated that absence of either FAS1 or FAS2 would render the CAF-1 complex non-functional, and therefore expected similar phenotypes for the two sets of crosses. However, all plants carrying the *fas1-4* allele show less severe defects, suggesting that *fas1-4* is not a complete loss-of-function mutant. Indeed, while RT-PCR analysis confirmed the absence of *FAS2* full-length transcripts in *fas2-5* mutants (Figure S5j), remaining full-length *FAS1* transcripts were detected in *fas1-4* mutants (Figure S5k,l) (Ramirez-Parra and Gutierrez, 2007). Therefore, *fas1-4 hira-1* plants retained residual CAF-1 activity, permitting survival and flowering.

Table 1 Epistatic relationship between CAF-1 and HIR complexes

Genotype		+/+	+/−	−/−	*n*	*N*
fas1-4	*cabin1-2*	17	27	19	63	3
	ubn2-2	28	28	3	59	3
	ubn1-2	22	27	9	58	3
	hira-1[b]	33	76	3	112	[a]
fas2-5	*cabin1-2*[b]	18	68	19	105	3
	ubn2-2	27	44	14	85	4
	ubn1-2[b]	22	73	19	114	3
	hira-1[b]	10	61	0	71	4

Number of F_2 plants with the indicated genotype. F_2 plants with the serrated leaf phenotype associated with the *fas* mutation were selected from the progeny of several independent F_1 plants, and genotyped for the corresponding HIR complex mutation (+/+, WT; +/−, heterozygous; −/−, homozygous for the studied mutation). The expected segregation ratio of WT to heterozygous to homozygous is 1:2:1 for genetically non-linked mutations. *n*, total number of plants analyzed; *N*, number of independent F_1 plants used in this study. The genes *FAS1*, *UBN2* and *UBN1* are located on the same chromosome (Figure S5a).
[a]Segregation analysis from a pool of seeds derived from several independent F_1 plants.
[b]Segregation significantly different from the expected ratio (chi-square test).

We conclude that the simultaneous mutation of CAF-1 and HIR complexes causes strong defects ranging from severe growth and developmental difficulties to lethality. While UBN1 appears to be dispensable for plant survival and reproduction, loss of HIRA causes the severest morphological aberrations, suggesting that HIRA is essential for HIR complex function. We assessed this hypothesis by generating *hira-1 ubn2-2*, *hira-1 ubn1-2* and *hira-1 cabin1-2* double mutants. We found that the double mutants show seed set similar to that of single *hira-1* mutant plants, with slightly more aborted and fewer viable seeds for *hira-1 ubn2-2* and *hira-1 cabin1-2* double mutants, respectively (Figure 5h). Furthermore, we observed comparable TSI levels in *hira-1*, *hira-1 ubn1-2* and *hira-1 cabin1-2* mutants, while *hira-1 ubn2-2* plants showed stronger reactivation of TSI repeat expression (Figure 5i). These observations, together with our data derived from crosses with mutants of the CAF-1 complex, suggest a central role of HIRA in HIR complex function, but do not exclude HIR complex-independent roles for the various subunits of the complex.

CAF-1 and HIR complexes are involved in independent but complementary pathways of chromatin assembly

The *fas1-4 hira-1* plants offer the unique opportunity to study the molecular consequences of simultaneous mutation of CAF-1 and HIR complexes. To obtain sufficient plant material, we selected double mutants grown on soil by phenotype from the segregating progeny of *fas1-4/FAS1 hira-1* plants (Figure S6a). We first analyzed H3 levels in *fas1-4 hira-1* plants. We found that the extractable pool of H3 proteins is significantly reduced compared to WT (Figure 6a). Histone H3 proteins are stored and transported by ASF1 proteins that donate histones to the chromatin complexes mediating assembly. The balance of *ASF1* expression is altered in *fas1-4 hira-1* mutants compared to WT: *ASF1A* expression is down-regulated while that of *ASF1B* is up-regulated (Figure 6b). Such a change in ASF1A and B protein levels may adjust histone flow to the remaining assembly complexes.

Given that *fas1-4 hira-1* plants are severely affected in growth and development, we then assessed whether heterochromatin organization and silencing are impaired in the double mutants. RT-PCR analysis on 4-week-old soil-grown plants showed that *fas1-4 hira-1* mutants release *TSI*, *Ta3* and *Mule* silencing to similar levels as single

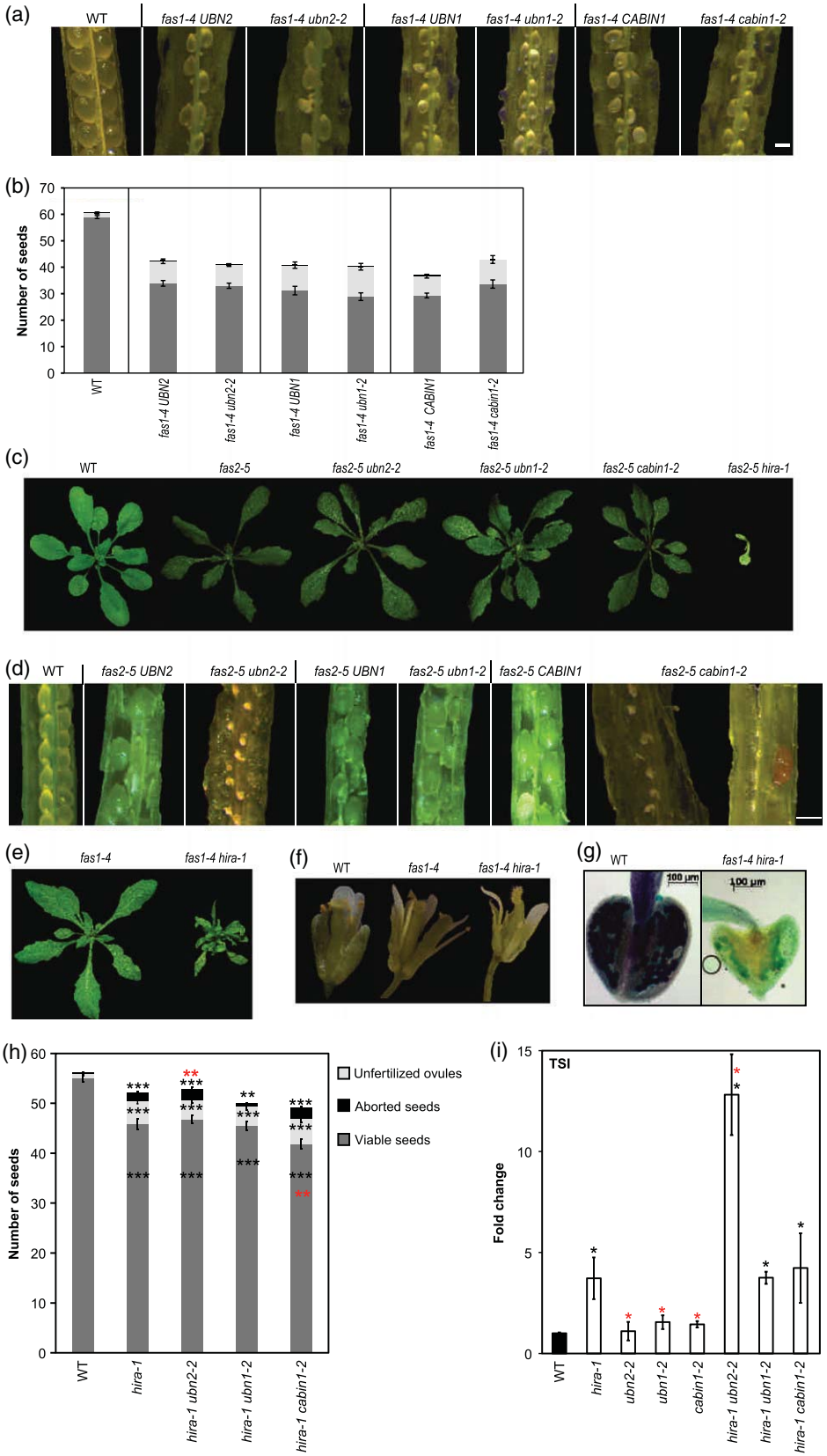


Figure 5. Genetic interactions between members of the CAF-1 and HIR complexes.
(a,d) Dissected siliques from F_2 progeny of *fas1-4* (a) and *fas2-5* (d) crosses with *ubn2-2*, *ubn1-2*, and *cabin1-2* mutants. For each cross, a representative silique from the *fas* single mutant and double mutant sister plants is shown. Scale bars = 0.3 mm.
(b) Quantification of seed content in siliques of *fas1-4* single and double mutants compared to WT. Thirty siliques pooled from at least four F_3 plants derived from F_2 plants homozygous for *fas1-4* and heterozygous for the respective mutation of the HIR complex were analyzed. No statistically significant difference was detected in comparison with the respective *fas1-4* sister plant (Student's *t* test).
(c,e) Representative WT, *fas2-5*, *fas2-5 ubn2-2*, *fas2-5 ubn1-1*, *fas2-5 cabin1-2 and fas2-5 hira-1* (c) and *fas1-4* and *fas1-4 hira-1* (e) mutant plants grown on soil.
(f) Representative flowers from WT, *fas1-4* and *fas1-4 hira-1* mutants.
(g) Representative WT and *fas1-4 hira-1* mutant anthers after Alexander staining, revealing absence of viable pollen in the double mutant.
(h) Quantification of seed content in WT, *hira-1*, *hira-1 ubn2-2*, *hira-1 ubn1-2* and *hira-1 cabin1-2* siliques. Quantifications were obtained from 30 pooled siliques from at least four plants. Red and black asterisks indicate comparisons with *hira-1* or WT, respectively.
(i) Quantitative RT-PCR analysis of TSI expression in WT and HIR complex single and double mutants. Red and black asterisks indicate comparisons with *hira-1* or WT, respectively.

mutants. Silencing at 106B and 180 bp repeats is not affected (Figure 6c,d). Furthermore, fluorescence *in situ* hybridization revealed that the global organization of 180 bp and TSI repetitive elements into chromocenters is not altered in *fas1-4 hira-1* nuclei (Figure S6b). Therefore, our data indicate that organization and silencing of repetitive sequences is largely maintained despite altered histone flow in the presence of CAF-1 and HIR complex mutations, suggesting that alternative silencing mechanisms counterbalance impaired nucleosome assembly in plants.

To investigate whether histone deposition mediated by the HIR complex rescues deficient replication-dependent nucleosome assembly in CAF-1 mutants, we assessed nucleosome occupancy by H3-ChIP combined with quantitative PCR on 4-week-old soil-grown *fas1-4 hira-1* plants. We found that nucleosome occupancy was reduced at all targets tested, including heterochromatic repeats, Ta3 and *Mule* transposons and an intergenic region, as well as at most analyzed regions of the euchromatic loci tested (Figure 6e–g). We conclude that, when CAF-1 and HIR-mediated assembly pathways are simultaneously impaired, plants fail to maintain nucleosome occupancy in both actively transcribed and transcriptionally repressed genomic regions. This suggested that deficient CAF-1-mediated incorporation of canonical histones could be rescued by alternative pathways, such as HIR complex-mediated assembly. Given the molecular and phenotypic defects caused by loss of both chaperone complexes, we conclude that Arabidopsis CAF-1 and HIR complexes are involved in independent pathways of chromatin assembly that concomitantly contribute to maintenance of nucleosome occupancy.

DISCUSSION

Proper packaging of DNA into chromatin is essential for genome structure, and ensures stability and inheritance of epigenetic information. A particular role in these processes may be assigned to factors responsible for histone deposition. Indeed, the various chromatin assembly factors, as well as other histone chaperones involved in histone transport and storage, are highly conserved through evolution.

Distinct contribution of the various members of the HIR complex

Various studies (Phelps-Durr *et al.*, 2005; Ingouff *et al.*, 2010; Nie *et al.*, 2014) have reported developmental defects ranging from WT appearance to distinctive leaf and flower phenotypes, suggesting that different growth conditions may play a role in the penetrance of *hira* mutant phenotypes. Although vegetative development was barely affected in single mutants for the various subunits under our growth conditions, we found defects in reproductive development in *hira-1* plants. Fewer viable seeds were also observed in the $hira^{amiRNA}$ line and in out-crosses of the *hira-1* mutant, suggesting that this defect is not caused by an epimutation. The simultaneous occurrence of fewer viable seeds, and an increased number of unfertilized ovules and aborted seeds in *hira* mutants, suggests a complex phenotype that is likely to have both gametophytic and sporophytic origin, and may result from alterations in the transcriptome (Nie *et al.*, 2014). Another intriguing hypothesis is the implication of HIRA in reprogramming of the histone variant repertoire (Ingouff *et al.*, 2007, 2010), as essentially only H3.3 and H3.3-like variants are present in sperm and egg cells (Ingouff *et al.*, 2010).

We revealed various levels of importance for the distinct subunits of the complex, with strongest defects observed in *hira-1* mutants. Given that HIRA comprises the conserved Hira domain and WD40 repeats involved in interactions with CABIN1 and UBN1 in mammals (Rai *et al.*, 2011; Tang *et al.*, 2012), and that Arabidopsis HIRA mediates the interaction with UBN2 and ASF1 (Nie *et al.*, 2014), its absence may destabilize the complex. This is the case in yeast, where loss of Hir1 or Hir2 reduces Hir3 protein levels (Song *et al.*, 2013), and in mammals, where HIRA depletion leads to a concomitant decrease in UBN1 and CABIN1 (Ray-Gallet *et al.*, 2011). However, our data provide evidence that the Arabidopsis HIR complex is at least still partly functional in the absence of UBN2 or CABIN1, and can sustain vegetative development in the absence of CAF-1. This reflects the situation in mammals, in which CABIN1 plays only a limited role in H3.3 deposition, followed by UBN1 and finally HIRA, which is crucial for H3.3

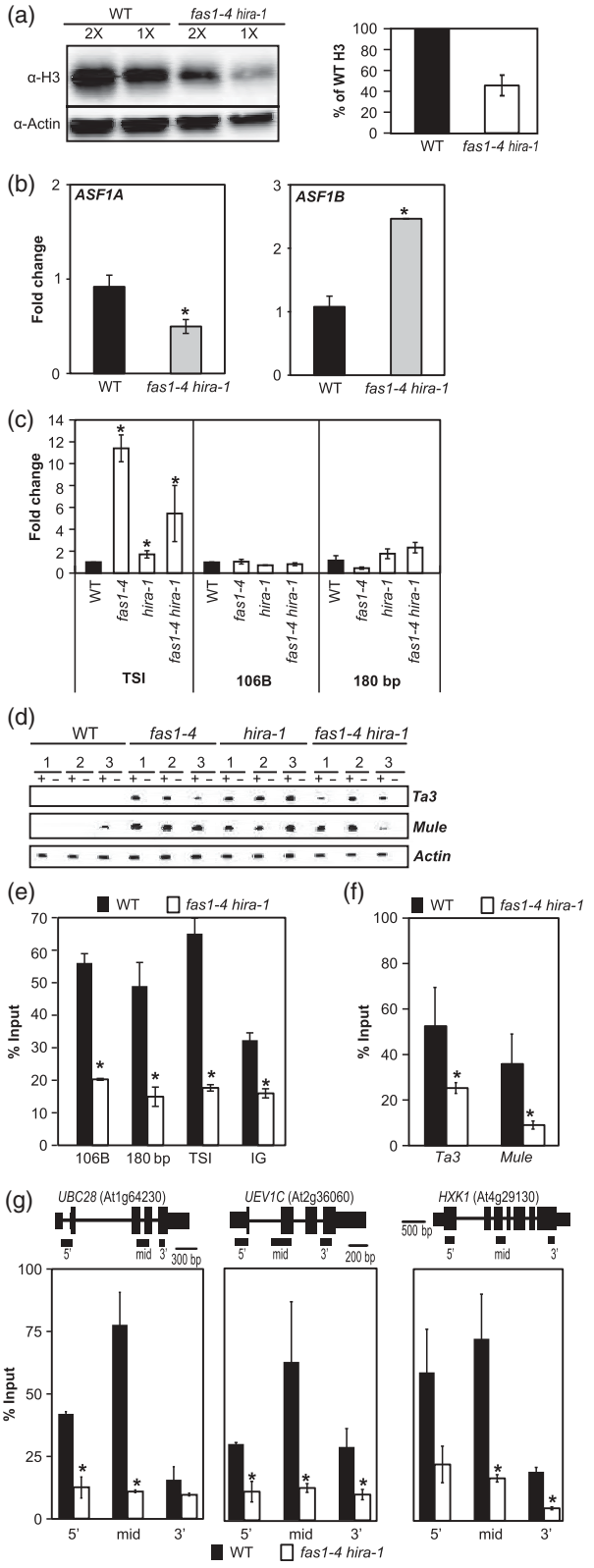


Figure 6. Molecular consequences of simultaneous mutation of CAF and HIR complexes.
(a) Left: Histone H3 protein levels quantified by Western blotting. Proteins extracted from WT and *fas1-4 hira-1* 4-week-old mutant shoots were loaded in each lane (2X, 20 μg; 1X, 10 μg). Right: Quantification of H3 band intensities relatively to actin from two independent experiments.
(b) Quantitative RT-PCR analysis of *ASF1A* and *ASF1B* expression in *fas1-4 hira-1* 4-week-old mutants.
(c) Quantitative RT-PCR analysis of TSI, 106B and 180 bp expression in WT, *fas1-4*, *hira-1* and *fas1-4 hira-1* 4-week-old plants.
(d) RT-PCR analysis of *Ta3* and *Mule* (At2g15810) transcript levels in WT, *fas1-4*, *hira-1* and *fas1-4 hira-1* 4-week-old plants. Three independent biological replicates were analyzed. *Actin* was used as a control.
(e–g) Histone H3 occupancy assessed by H3-ChIP combined with quantitative PCR relative to input in WT and *fas1-4 hira-1* 4-week-old mutant shoots at heterochromatic and intergenic regions (e,f) and at various positions along three active genes (g).

incorporation (Ray-Gallet *et al.*, 2011). Furthermore, transcriptional repression of yeast histone genes depends on Hir1 and Hir2, but only partly on Hir3 (Osley and Lycan, 1987; Spector *et al.*, 1997), and the *Drosophila* HIR complex does not comprise a CABIN1 ortholog (Amin *et al.*, 2011). An alternative but not exclusive hypothesis is that HIRA and other subunits of the complex also have HIR complex-independent functions. Indeed, genome-wide ChIP-Seq analysis in human cells localized HIRA to several chromosomal positions that are not co-occupied by UBN1 or ASF1 (Pchelintsev *et al.*, 2013), and roles as transcriptional co-repressors have been demonstrated for UBN1 and CABIN1 (Aho *et al.*, 2000; Jang *et al.*, 2007). We expect that specific functions of the various Arabidopsis HIR complex subunits may be revealed only under certain growth and environmental conditions, or in particular cell types.

Role of HIRA in nucleosome dynamics

Loss of HIRA or FAS1 subunits results in reduced levels of an extractable histone pool. In comparison to methods developed for mammalian cell cultures (Loyola *et al.*, 2006), this fraction of extractable histones probably presents a combination of the cytosolic and nucleoplasmic pool of histones, while nucleosomal histones were precipitated during the centrifugation step. Therefore, the reduced levels of this histone pool in *fas1-4* and *hira-1* mutants suggest that absence of histone deposition complexes affects histone flow, and may render prone to degradation those histones that are not readily assembled into chromatin. This alteration in the extractable histone pool is then reflected at the level of nucleosomal H3 occupancy. Indeed, by analyzing H3 enrichment in a locus-specific manner, we found that loss of HIRA affects nucleosome occupancy at certain active genes, specifically within the body and 3′ ends, regions that were previously shown to be enriched in H3.3 (Stroud *et al.*, 2012; Wollmann *et al.*, 2012). This nucleosome loss at active genes may be explained by failure to restore nucleosomes using H3.3 after passage of the transcriptional machinery. Interestingly, similar to CAF-1, HIRA also affects H3 occupancy at heterochromatic targets. Despite heterochromatin being not (or very poorly) transcribed, and generally considered to have slow histone

exchange rates, the presence of H3.3 has been reported in centric and pericentric heterochromatin and telomeres in mouse and human (Wong *et al.*, 2009; Drané *et al.*, 2010; Goldberg *et al.*, 2010; Lewis *et al.*, 2010; Dunleavy *et al.*, 2011; Morozov *et al.*, 2012). It is therefore possible that HIRA is implicated in histone deposition at heterochromatic sites, a process that may be particularly important in cells that undergo neither replication nor endoreplication. Such a requirement for the HIR complex may explain the transcriptional reactivation of endogenous pericentromeric repeats and transposons observed in *hira-1* mutants, and implies that nucleosome occupancy maintenance contributes to transcriptional gene silencing. Interestingly, HIRA orthologs in fission yeast are also involved in transcriptional silencing at pericentromeric heterochromatin (Blackwell and Martin, 2004; Yamane *et al.*, 2011). Further studies that take the cellular properties of various tissues and developmental variations into account should shed further light on cell-specific roles of the HIR complex.

Several lines of evidence have linked HIRA and deposition of histone replacement variants to transcription control (Ahmad and Henikoff, 2002; Formosa *et al.*, 2002; Sakai *et al.*, 2009; Ray-Gallet *et al.*, 2011; Schneiderman *et al.*, 2012). However, when we analyzed the capacity of HIRA-deficient plants to activate or repress gene transcription in response to salt stress, we found no difference compared to WT, either for genes with basal expression levels or for genes that were silent before induction. To obtain a more general view regarding the role of HIRA in the transcriptional response to environmental stimuli, additional salt-responsive genes should be analyzed and different abiotic stresses considered in the future. Nevertheless, based on the available data, we speculate that either HIR-mediated histone dynamics or H3.3 incorporation during transcription are dispensable for proper expression at these sites. The possibility that H3.3 incorporation is dispensable is in agreement with observations from both *Tetrahymena* and *Drosophila*, which survive and show correct gene expression in the absence of H3.3 (Cui *et al.*, 2006; Hödl and Basler, 2009), and the lack of strong phenotypes in *hira-1*. It is likely that alternative histone chaperone complexes are also implicated in deposition of replacement variants. While no DAXX ortholog has been identified in plants (Zhu *et al.*, 2013), Arabidopsis encodes several DEK proteins (Waidmann *et al.*, 2014), and other as yet unidentified histone chaperones may play a role in histone variant dynamics (Otero *et al.*, 2014).

Functional redundancy of CAF-1 and HIR complexes in histone deposition?

The combination of the *fas2-5* and the *hira-1* mutations causes post-germination lethality, while *fas1-4 hira-1* mutants are viable but sterile. The survival of *fas1-4 hira-1* mutants may be explained by residual CAF-1 activity, concomitant with the slow growth and potentially the premature switch to endoreplication in *fas* mutants (Ramirez-Parra and Gutierrez, 2007). In addition, the shuttling of histones primarily to the CAF-1 complex by changing the ASF1A/B balance may help to sustain sufficient nucleosome assembly to permit plant survival. Maintenance of heterochromatin in the absence of CAF-1 function in yeast requires the HIR complex (Osley and Lycan, 1987; Kaufman *et al.*, 1998). However, despite a reduction in H3 occupancy in the *fas1-4 hira-1* double mutant, silencing of an endogenous repeat sequence and certain transposons as well as heterochromatin organization is not further affected compared to the single mutants. This is surprising, as appropriate nucleosome assembly is not only required for DNA packaging but is also important for propagating epigenetic information. We interpret this finding as due to the presence of alternative mechanisms required for gene silencing and heterochromatin maintenance, such as DNA methylation, which is of primary importance in plants in directing histone modifications and silencing (Rigal and Mathieu, 2011), and which is unaffected in *fas* mutants (Schönrock *et al.*, 2006).

The nucleosomal loss and accentuated phenotypic defects in *fas1-4 hira-1* double mutants, together with the viability of the respective single mutants, suggest partial functional redundancy between the histone deposition complexes. Hence, we speculate that the Arabidopsis HIR complex can at least partially rescue a deficiency in replication-coupled histone assembly, potentially by nucleosomal gap-filling mechanisms. Such a mechanism has been proposed in mammals, where, in the absence of functional CAF-1, the HIR complex recognizes naked DNA stretches remaining after replication via its DNA-binding properties, and fills the gaps through H3 deposition (Ray-Gallet *et al.*, 2011). Whether this mechanism occurs in plants remains to be determined. Given the moderate nucleosomal reduction observed in *hira-1* mutants, we envisage the occurrence of histone deposition by other complexes dedicated to replacement variant assembly or even H3.3 incorporation by the CAF-1 complex. Indeed, in DAXX- or HIRA-depleted mammalian cells, CAF-1 co-purifies with H3.3 (Drané *et al.*, 2010; Lewis *et al.*, 2010).

In conclusion, several lines of evidence support the notion that the Arabidopsis HIR complex functions as a histone chaperone: loss of HIRA reduces the pool of extractable H3 histones, affects nucleosome occupancy at both active and repressed chromosomal regions, and alleviates transcriptional silencing. In addition, synthetic lethality is observed when the *hira* mutant is combined with complete loss-of-function *fas* alleles. Taken together, our results imply an evolutionarily conserved mode of action of these two plant histone H3 assembly complexes. Indeed, the presence of Arabidopsis H3.1 and H3.3 variants at

transcriptionally repressed or active genomic regions, respectively, is similar to *Drosophila* and mammals (Mito *et al.*, 2005; Goldberg *et al.*, 2010; Stroud *et al.*, 2012; Wollmann *et al.*, 2012; Shu *et al.*, 2014). This observation was unexpected, given that the evolution of functionally divergent H3 variants is thought to have occurred independently in plant and animal kingdoms. However, the highly conserved function of plant histone chaperone complexes suggests that histone chaperones may have contributed to functional diversification of the histones they transport and deposit.

EXPERIMENTAL PROCEDURES

Plant material

Mutant Arabidopsis lines were obtained from the Nottingham Arabidopsis Stock Center (http://arabidopsis.info/) and/or were gifts from other laboratories. Homozygous mutants *fas1-4* (SAIL-662-D10) (Kirik *et al.*, 2006), *fas2-5* (SALK_147693), *hira-1* (WiscDsLox362H05) (Ingouff *et al.*, 2010), *ubn2-2* (GABI_018D02), *ubn1-2* (GABI_130H01), *cabin1-2* (SALK_099927), SALK_019573 (Phelps-Durr *et al.*, 2005), GABI_775H03 and SALK_143806 (Phelps-Durr *et al.*, 2005) were identified by PCR-based genotyping (Table S2, Data S1). All mutants are in the Columbia background. Plants were grown on soil in a growth chamber under 16 h light/8 h dark cycles at 22°C. For *in vitro* culture, seeds were sterilized and sown on germination medium containing 0.8% w/v agar, 1% w/v sucrose and Murashige & Skoog salts (M0255; Duchefa Biochemie; www.duchefa-biochemie.com). After 2 days of stratification at 4°C at dark, plants were grown under 16 h light/8 h dark cycles at 23°C.

Salt treatment

Plantlets grown for 10 days *in vitro* were transferred into liquid MS medium containing 200 mM NaCl. Samples were shock-frozen in liquid nitrogen before treatment and after 1 h of salt exposure.

RNA extraction and RT-PCR

RNA was extracted using Tri-Reagent (Euromedex; www.euromedex.com) according to the manufacturer's instructions, treated with RQ1 DNase I (Promega; www.promega.com), and purified using phenol/chloroform extraction. Reverse transcription was performed using either oligo$(dT)_{15}$ or random hexamers and M-MLV reverse transcriptase (Promega). The resulting cDNAs were used in standard PCR (Promega Flexi) or in quantitative PCR using the LightCycler® 480 SYBR Green I master kit on the Roche LightCycler® 480 (lifescience.roche.com). Transcript levels of interest were normalized to those for At2g28390 (Czechowski *et al.*, 2005) using the comparative threshold cycle method. Quantitative RT-PCR histograms show means of transcript levels ± SEM obtained for two independent PCR amplifications of three biological replicates. The *y* axis shows the fold change relative to WT (WT set to 1) after normalization to expression of At2g28390.

Statistical analysis

Student's *t* test was used for mean comparison for quantitative PCR analyses and seed counting: Asterisks in the figures indicate statistically significant differences (***$P < 0.001$; **$P < 0.01$; *$P < 0.05$).

Ploidy analysis

Nuclei were prepared using a modified version of the original Galbraith method (Galbraith *et al.*, 1983). For ploidy analysis of each genotype, two independent preparations of pooled shoot material from 15 plants were investigated. Cytometric analysis was performed using an Attune® acoustic focusing cytometer (Thermo Fisher Scientific; www.lifetechnologies.com). For statistical analysis of differences in ploidy levels, we performed a Shapiro–Wilk test, followed by a Bartlett's test, and finally Tukey's HSD (honest significant difference) test was applied in conjunction with ANOVA to find means that were significantly different (indicated by an asterisk).

ChIP analysis

Chromatin of plantlets was formaldehyde cross-linked, and chromatin immunoprecipitation was performed as previously described (Bowler *et al.*, 2004) with minor modifications. Chromatin was sheared using a Diagenode bioruptor (ten cycles of 30 sec on and 1.5 min off). Protein A-coupled magnetic beads (Diagenode; www.diagenode.com) or Protein A-coupled Dynabeads (Invitrogen) were used, and the sonicated chromatin was pre-cleared in presence of magnetic beads for 3 h, before immunoprecipitation with anti-H3 antibody (Abcam, ab1791; www.abcam.com). DNA was quantified using quantitative PCR (Roche) and normalized relative to input.

Protein extraction and western blotting

Proteins were extracted from 100 mg of plantlet material as described previously (Durut *et al.*, 2014). Western blots were probed using anti-H3 antibody (Abcam, ab1791, 1/5000). Equal loading was confirmed using anti-actin antibody (1/1000; Sigma; www.sigmaaldrich.com) and Ponceau staining. Primary antibodies were revealed by incubation with anti-rabbit (1/25 000; Sigma) or anti-mouse (1/5000; Sigma) secondary antibodies. Immunoblot analysis was performed using ECL Western blotting detection reagents (GE Healthcare Bio-Sciences; www.gelifesciences.com). Densitometric analysis of immunoreactive protein bands was performed on non-saturated signals using MultiGauge software (Fujifilm; www.fujifilm.com), and H3 levels normalized to actin in WT were set to 100%.

ACKNOWLEDGEMENTS

C.D., S.L.G. and A.V.P. were supported by Agence Nationale de la Recherche grant 'Dynam'Het' ANR-11 JSV2 009 01, and S.C. and A.V.P. were supported by Agence Nationale de la Recherche grant 'SINODYN' ANR-12 ISV6 0001. M.B., A.P. and L.S. are supported by stipends from the Region Auvergne and M.B. is supported by an Association pour la Recherche sur le Cancer doctoral fellowship. This work was supported by the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Université Blaise Pascal and the Université d'Auvergne. We thank M.-C. Espagnol and L. Borao for technical assistance, S. Tourmente for stimulating discussion, and O. Mittelsten Scheid and J. Bassler (Gregor Mendel Institute, Vienna) for critical comments on the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Evolutionary conservation and expression of the HIR complex subunits.

Figure S2. Characterization of HIR and CAF-1 complex mutants.

Figure S3. Enrichment in H3.1 and H3.3 at targets analyzed by H3-ChIP combined with quantitative PCR.

Figure S4. Analysis of transcriptional silencing release in CAF-1 and HIR complex mutants.

Figure S5. Genetic interactions between members of the CAF-1 and HIR complexes.

Figure S6. Generation and study of *fas1-4 hira-1* double mutant material.

Table S1. *fas hira-1* double mutants are under-represented in segregating F_2 populations.

Table S2. List of primers used in the study.

Data S1. Supplementary experimental procedures.

REFERENCES

Ahmad, K. and Henikoff, S. (2002) The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9**, 1191–1200.

Aho, S., Buisson, M., Pajunen, T., Ryoo, Y.W., Giot, J.F., Gruffat, H., Sergeant, A. and Uitto, J. (2000) Ubinuclein, a novel nuclear protein interacting with cellular and viral transcription factors. *J. Cell Biol.* **148**, 1165–1176.

Amin, A.D., Vishnoi, N. and Prochasson, P. (2011) A global requirement for the HIR complex in the assembly of chromatin. *Biochim. Biophys. Acta*, **1819**, 264–276.

Balaji, S., Iyer, L.M. and Aravind, L. (2009) HPC2 and ubinuclein define a novel family of histone chaperones conserved throughout eukaryotes. *Mol. BioSyst.* **5**, 269–275.

Banumathy, G., Somaiah, N., Zhang, R. *et al.* (2009) Human UBN1 is an ortholog of yeast Hpc2p and has an essential role in the HIRA/ASF1a chromatin-remodeling pathway in senescent cells. *Mol. Cell. Biol.* **29**, 758–770.

Blackwell, C. and Martin, K. (2004) Schizosaccharomyces pombe HIRA-like protein Hip1 is required for the periodic expression of histone genes and contributes to the function of complex centromeres. *Mol. Cell. Biol.* **24**, 4309–4320.

Bonnefoy, E., Orsi, G.A., Couble, P. and Loppin, B. (2007) The essential role of *Drosophila* HIRA for *de novo* assembly of paternal chromatin at fertilization. *PLoS Genet.* **3**, 1991–2006.

Bowler, C., Benvenuto, G., Laflamme, P., Molino, D., Probst, A.V., Tariq, M. and Paszkowski, J. (2004) Chromatin techniques for plant cells. *Plant J.* **39**, 776–789.

Campos, E.I., Fillingham, J., Li, G. *et al.* (2010) The program for processing newly synthesized histones H3.1 and H4. *Nat. Struct. Mol. Biol.* **17**, 1343–1351.

Cui, B., Liu, Y. and Gorovsky, M.A. (2006) Deposition and function of histone H3 variants in *Tetrahymena thermophila*. *Mol. Cell. Biol.* **26**, 7719–7730.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K. and Scheible, W.R. (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol.* **139**, 5–17.

De Koning, L., Corpet, A., Haber, J.E. and Almouzni, G. (2007) Histone chaperones: an escort network regulating histone traffic. *Nat. Struct. Mol. Biol.* **14**, 997–1007.

Drané, P., Ouararhni, K., Depaux, A., Shuaib, M. and Hamiche, A. (2010) The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev.* **24**, 1253–1265.

Duc, C., Sherstnev, A., Cole, C., Barton, G.J. and Simpson, G.G. (2013) Transcription termination and chimeric RNA formation controlled by *Arabidopsis thaliana* FPA D. C. Baulcombe, ed. *PLoS Genet.* **9**, e1003867.

Dunleavy, E.M., Almouzni, G. and Karpen, G.H. (2011) H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G phase. *Nucleus*, **2**, 146–157.

Durut, N., Abou-Ellail, M., Pontvianne, F. *et al.* (2014) A duplicated NUCLEOLIN gene with antagonistic activity is required for chromatin organization of silent 45S rDNA in Arabidopsis. *Plant Cell*, **1**, 1–16.

English, C.M., Adkins, M.W., Carson, J.J., Churchill, M.E.A. and Tyler, J.K. (2006) Structural basis for the histone chaperone activity of Asf1. *Cell*, **127**, 495–508.

Exner, V., Taranto, P., Schönrock, N., Gruissem, W. and Hennig, L. (2006) Chromatin assembly factor CAF–1 is required for cellular differentiation during plant development. *Development*, **133**, 4163–4172.

Filipescu, D., Szenker, E. and Almouzni, G. (2013) Developmental roles of histone H3 variants and their chaperones. *Trends Genet.* **29**, 630–640.

Formosa, T., Ruone, S. and Adams, M. (2002) Defects in SPT16 or POB3 (yFACT) in *Saccharomyces cerevisiae* cause dependence on the Hir/Hpc pathway: polymerase passage may degrade chromatin structure. *Genetics*, **3**, 1557–1571.

Gaillard, P.H., Martini, E.M., Kaufman, P.D., Stillman, B., Moustacchi, E. and Almouzni, G. (1996) Chromatin assembly coupled to DNA repair: a new role for chromatin assembly factor I. *Cell*, **86**, 887–896.

Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P. and Firoozabady, E. (1983) Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science*, **220**, 1049–1051.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M. *et al.* (2010) Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell*, **140**, 678–691.

Groth, A., Rocha, W., Verreault, A. and Almouzni, G. (2007) Chromatin challenges during DNA replication and repair. *Cell*, **218**, 721–7331.

Hödl, M. and Basler, K. (2009) Transcription in the absence of histone H3.3. *Curr. Biol.* **19**, 1221–1226.

Houlard, M., Berlivet, S., Probst, A.V., Quivy, J.-P.P., Héry, P., Almouzni, G. and Gérard, M. (2006) CAF-1 is essential for heterochromatin organization in pluripotent embryonic cells. *PLoS Genet.* **2**, e181.

Ingouff, M. and Berger, F. (2010) Histone3 variants in plants. *Chromosoma*, **119**, 27–33.

Ingouff, M., Hamamura, Y., Gourgues, M., Higashiyama, T. and Berger, F. (2007) Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. *Curr. Biol.* **17**, 1032–1037.

Ingouff, M., Rademacher, S., Holec, S. *et al.* (2010) Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in Arabidopsis. *Curr. Biol.* **20**, 2137–2143.

Jang, H., Choi, D.-E., Kim, H., Cho, E.-J. and Youn, H.-D. (2007) Cabin1 represses MEF2 transcriptional activity by association with a methyltransferase, SUV39H1. *J. Biol. Chem.* **282**, 11172–11179.

Jin, C. and Felsenfeld, G. (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev.* **21**, 1519–1529.

Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K. and Felsenfeld, G. (2009) H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat. Genet.* **41**, 941–945.

Kaufman, P.D., Cohen, J.L. and Osley, M. (1998) Hir proteins are required for position-dependent gene silencing in *Saccharomyces cerevisiae* in the absence of chromatin assembly factor I. *Mol. Cell. Biol.* **18**, 4793–4806.

Kaya, H., Shibahara, K.I., Taoka, K.I., Iwabuchi, M., Stillman, B. and Araki, T. (2001) FASCIATA genes for chromatin assembly factor-1 in arabidopsis maintain the cellular organization of apical meristems. *Cell*, **104**, 131–142.

Kirik, A., Pecinka, A., Wendeler, E. and Reiss, B. (2006) The chromatin assembly factor subunit FASCIATA1 is involved in homologous recombination in plants. *Plant Cell*, **18**, 2431–2442.

Klapholz, B., Dietrich, B.H., Schaffner, C., Hérédia, F., Quivy, J.-P., Almouzni, G. and Dostatni, N. (2009) CAF-1 is required for efficient replication of euchromatic DNA in *Drosophila* larval endocycling cells. *Chromosoma*, **118**, 235–248.

Lewis, P.W., Elsaesser, S.J., Noh, K.M., Stadler, S.C. and Allis, C.D. (2010) Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc. Natl Acad. Sci. USA* **107**, 14075–14080.

Loppin, B., Bonnefoy, E., Anselme, C., Laurençon, A., Karr, T.L. and Couble, P. (2005) The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. *Nature*, **437**, 1386–1390.

Loyola, A., Bonaldi, T., Roche, D., Imhof, A. and Almouzni, G. (2006) PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. *Mol. Cell*, **24**, 309–316.

Mito, Y., Henikoff, J.G. and Henikoff, S. (2005) Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* **37**, 1090–1097.

Morel, J.B., Mourrain, P., Béclin, C. and Vaucheret, H. (2000) DNA methylation and chromatin structure affect transcriptional and post-transcriptional transgene silencing in Arabidopsis. *Curr. Biol.* **10**, 1591–1594.

Morozov, V., Gavrilova, E., Ogryzko, V. and Ishov, A. (2012) Dualistic function of Daxx at centromeric and pericentromeric heterochromatin in normal and stress conditions. *Nucleus*, **3**, 276–285.

Natsume, R., Eitoku, M., Akai, Y. and Sano, N. (2007) Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. *Nature*, **446**, 1–4.

Nie, X., Wang, H., Li, J., Holec, S. and Berger, F. (2014) The HIRA complex that deposits the histone H3.3 is conserved in Arabidopsis and facilitates transcriptional dynamics. *Biol. Open*, **3**, 794–802.

Ono, T., Kaya, H., Takeda, S. *et al.* (2006) Chromatin assembly factor 1 ensures the stable maintenance of silent chromatin states in Arabidopsis. *Genes Cells*, **11**, 153–162.

Osley, M.A. and Lycan, D. (1987) Trans-acting regulatory mutations that alter transcription of *Saccharomyces cerevisiae* histone genes. *Mol. Cell. Biol.* **7**, 4204–4210.

Otero, S., Desvoyes, B. and Gutierrez, C. (2014) Histone h3 dynamics in plant cell cycle and development. *Cytogenet. Genome Res.* **143**, 114–124.

Pchelintsev, N.A., Mcbryan, T., Rai, T.S., van Tuyn, J., Ray-Gallet, D., Almouzni, G. and Adams, P.D. (2013) Placing the HIRA histone chaperone complex in the chromatin landscape. *Cell Rep.* **3**, 1012–1019.

Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N. and Mittelsten Scheid, O. (2010) Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *Plant Cell*, **22**, 3118–3129.

Petesch, S. and Lis, J. (2012) Overcoming the nucleosome barrier during transcript elongation. *Trends Genet.* **28**, 285–294.

Phelps-Durr, T.L., Thomas, J., Vahab, P. and Timmermans, M.C.P. (2005) Maize rough sheath2 and its Arabidopsis orthologue ASYMMETRIC LEAVES1 interact with HIRA, a predicted histone chaperone, to maintain knox gene silencing and determinacy during organogenesis. *Plant Cell*, **17**, 2886–2898.

Rai, T.S., Puri, A., McBryan, T. *et al.* (2011) Human CABIN1 is a functional member of the human HIRA/UBN1/ASF1a histone H3.3 chaperone complex. *Mol. Cell. Biol.* **31**, 4107–4118.

Ramirez-Parra, E. and Gutierrez, C. (2007) E2F regulates FASCIATA1, a chromatin assembly gene whose loss switches on the endocycle and activates gene expression by changing the epigenetic status. *Plant Physiol.* **144**, 105–120.

Ray-Gallet, D., Quivy, J.P., Scamps, C., Martini, E.M., Lipinski, M. and Almouzni, G. (2002) HIRA is critical for a nucleosome assembly pathway independent of DNA synthesis. *Mol. Cell*, **9**, 1091–1100.

Ray-Gallet, D., Woolfe, A., Vassias, I. *et al.* (2011) Dynamics of histone H3 deposition *in vivo* reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. *Mol. Cell*, **44**, 928–941.

Rigal, M. and Mathieu, O. (2011) A 'mille-feuille' of silencing: Epigenetic control of transposable elements. *Biochim. Biophys. Acta*, **1809**, 452–458.

Roberts, C., Sutherland, H.F., Farmer, H., Kimber, W., Halford, S., Carey, A., Brickman, J.M., Wynshaw-Boris, A. and Scambler, P.J. (2002) Targeted mutagenesis of the Hira gene results in gastrulation defects and patterning abnormalities of mesoendodermal derivatives prior to early embryonic lethality. *Mol. Cell. Biol.* **22**, 2318–2328.

Sakai, A., Schwartz, B.E., Goldstein, S. and Ahmad, K. (2009) Transcriptional and developmental functions of the H3.3 histone variant in *Drosophila*. *Curr. Biol.* **19**, 1816–1820.

Sawatsubashi, S., Murata, T., Lim, J. *et al.* (2010) A histone chaperone, DEK, transcriptionally coactivates a nuclear receptor. *Genes Dev.* **24**, 159–170.

Schneiderman, J.I., Orsi, G.A., Hughes, K.T., Loppin, B. and Ahmad, K. (2012) Nucleosome-depleted chromatin gaps recruit assembly factors for the H3.3 histone variant. *Proc. Natl Acad. Sci.* **109**, 19721–19726.

Schönrock, N., Exner, V., Probst, A.V., Gruissem, W. and Hennig, L. (2006) Functional genomic analysis of CAF-1 mutants in *Arabidopsis thaliana*. *J. Biol. Chem.* **281**, 9560–9568.

Schwartz, B.E. and Ahmad, K. (2005) Transcriptional activation triggers deposition and removal of the histone variant H3.3. *Genes Dev.* **19**, 804–814.

Shu, H., Nakamura, M., Siretskiy, A., Borghi, L., Moraes, I., Wildhaber, T., Gruissem, W. and Hennig, L. (2014) Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes. *Genome Biol.* **15**, R62.

Song, Y., Seol, J.H., Yang, J.H., Kim, H.J., Han, J.W., Youn, H.D. and Cho, E.J. (2013) Dissecting the roles of the histone chaperones reveals the evolutionary conserved mechanism of transcription-coupled deposition of H3.3. *Nucleic Acids Res.* **41**, 5199–5209.

Spector, M., Raff, A. and DeSilva, H. (1997) Hir1p and Hir2p function as transcriptional corepressors to regulate histone gene transcription in the *Saccharomyces cerevisiae* cell cycle. *Mol. Cell. Biol.* **17**, 545–552.

Steimer, A., Amedeo, P., Afsar, K., Fransz, P., Mittelsten Scheid, O. and Paszkowski, J. (2000) Endogenous targets of transcriptional gene silencing in Arabidopsis. *Plant Cell*, **12**, 1165–1178.

Stillman, B. (1989) Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication *in vitro*. *Cell*, **58**, 15–25.

Stroud, H., Otero, S., Desvoyes, B., Ramírez-Parra, E., Jacobsen, S.E. and Gutierrez, C. (2012) Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci.* **109**, 5370–5375.

Szenker, E., Lacoste, N. and Almouzni, G. (2012) A developmental requirement for HIRA-dependent H3.3 deposition revealed at gastrulation in Xenopus. *Cell Rep.* **1**, 730–740.

Tagami, H., Ray-Gallet, D., Almouzni, G. and Nakatani, Y. (2004) Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell*, **116**, 51–61.

Takeda, S., Tadele, Z., Hofmann, I. *et al.* (2004) BRU1, a novel link between responses to DNA damage and epigenetic gene silencing in Arabidopsis. *Genes Dev.* **18**, 782–793.

Talbert, P. and Henikoff, S. (2010) Histone variants – ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11**, 264–275.

Talbert, P.B., Ahmad, K., Almouzni, G. *et al.* (2012) A unified phylogeny-based nomenclature for histone variants. *Epigenetics Chromatin*, **5**, 7.

Tang, Y., Poustovoitov, M.V., Zhao, K., Garfinkel, M., Canutescu, A., Dunbrack, R., Adams, P.D. and Marmorstein, R. (2006) Structure of a human ASF1a-HIRA complex and insights into specificity of histone chaperone complex assembly. *Nat. Struct. Mol. Biol.* **13**, 921–929.

Tang, Y., Puri, A., Ricketts, M.D., Rai, T., Hoffmann, J., Hoi, E., Adams, P., Schultz, D. and Marmorstein, R. (2012) Identification of an Ubinuclein 1 region required for stability of the human HIRA/UBN1/CABIN1/ASF1a histone H3.3 chaperone complex. *Biochemistry*, **51**, 2366–2377.

Tyler, J. and Collins, K. (2001) Interaction between the *Drosophila* CAF–1 and ASF1 Chromatin Assembly Factors. *Mol. Cell. Biol.* **21**, 6574–6584.

Waidmann, S., Kusenda, B., Mayerhofer, J., Mechtler, K. and Jonak, C. (2014) A DEK domain-containing protein modulates chromatin structure and function in Arabidopsis. *Plant Cell*, **26**, 4328–4344.

Wollmann, H., Holec, S., Alden, K., Clarke, N., Jacques, P. and Berger, F. (2012) Dynamic deposition of histone variant h3.3 accompanies developmental remodeling of the Arabidopsis transcriptome. *PLoS Genet.* **8**, e1002658.

Wong, L.H., Ren, H., Williams, E. *et al.* (2009) Histone H3.3 incorporation provides a unique and functionally essential telomeric chromatin in embryonic stem cells. *Genome Res.* **19**, 404–414.

Yamane, K., Mizuguchi, T., Cui, B., Zofall, M., Noma, K. and Grewal, S.I. (2011) Asf1/HIRA facilitate global histone deacetylation and associate with HP1 to promote nucleosome occupancy at Heterochromatic loci. *Mol. Cell*, **41**, 56–66.

Yang, J.-H., Choi, J.-H., Jang, H., Park, J.-Y., Han, J.-W., Youn, H.-D. and Cho, E.-J. (2011) Histone chaperones cooperate to mediate Mef2-targeted transcriptional regulation during skeletal myogenesis. *Biochem. Biophys. Res. Commun.* **407**, 541–547.

Zeller, G., Henz, S., Widmer, C., Sachsenberg, T., Rätsch, G., Weigel, D. and Laubinger, S. (2009) Stress-induced changes in the *Arabidopsis thaliana* transcriptome analyzed using whole-genome tiling arrays. *Plant J.* **58**, 1068–1082.

Zhu, Y., Dong, A. and Shen, W.H. (2013) Histone variants and chromatin assembly in plant abiotic stress responses. *Biochim. Biophys. Acta*, **1819**, 343–348.

B. ATRX contrôle l'occupation nucléosomale et régule l'expression des gènes d'ARNr 45S

Dans l'objectif d'augmenter les connaissances sur la dynamique des histones chez *Arabidopsis thaliana*, nous avons caractérisé une autre protéine chaperonne : ATRX (Alpha Thalassemia-mental Retardation X-linked) (Shaked et al., 2006). Chez l'homme, ATRX dépose H3.3 au niveau des péricentromères et des télomères (Goldberg et al., 2010; Lewis et al., 2010).

Notre étude a montré que les mutants *atrx* d'*Arabidopsis thaliana* présentent des problèmes de développement et de fertilité. La perte d'ATRX entraine une diminution de la densité nucléosomale et une altération de la balance H3.3/H3.1 dans la cellule. De plus, la combinaison d'une mutation *atrx* et *hira* cause un phénotype sévère de développement, pendant que les mutants seuls montrent uniquement une fertilité légèrement réduite. Ceci suggère un rôle compensatoire entre HIR et ATRX dans le dépôt de H3.3. Nous avons aussi identifié ATRX comme un régulateur de la transcription de l'ARN 45S puisque les mutants *atrx* présentent une expression altérée des variants d'ARNr 45S comparé à une plante sauvage.

Similairement au travail concernant le complexe HIR, j'ai contribué à ce manuscrit par mon expertise en cytométrie en flux et j'ai participé à l'évaluation de la fertilité des plantes mutants ainsi que des croisement entre les différents mutants pour les sous-unités du complexe CAF-1 et d'ATRX par comptage des graines viables et avortées dans les siliques.

Arabidopsis ATRX controls nucleosome occupancy and regulates 45S rRNA expression

Céline Duc[1], Matthias Benoit[1], Gwénaëlle Détourné[1,2], Lauriane Simon[1], Axel Poulet[1,2], Samuel Le Goff[1], Sylviane Cotterell[1], Christophe Tatout[1], Aline V. Probst[1*]

[1] UMR CNRS 6293, Clermont Université, INSERM U1103, Génétique, Reproduction et Développement, 10 Avenue des Landais, TSA 60026, CS 60026, 63178 Aubière Cedex, France

[2] Department of Biological and Medical Sciences, Oxford Brookes University, Oxford OX3 0BP, UK.

[*] Corresponding author

Email: aline.probst@univ-bpclermont.fr.

Short title: AtATRX, a player in H3 incorporation landscape

Abstract

Histones are essential components of the nucleosome, the major chromatin subunit that structures linear DNA molecules and regulates access of other proteins to DNA. Specific histone chaperone complexes control the correct deposition of canonical and variant histones to modulate nucleosome structure and stability. In this study, we characterized Arabidopsis ATRX (Alpha Thalassemia-mental Retardation X-linked) using viable null alleles. The *atrx* mutants show defects in development and fertility as well as moderate deficiencies in DNA replication and repair. Genetic and molecular analyses reveal that ATRX plays a role in histone H3 deposition. ATRX loss-of-function affects cellular histone pools, changes the cellular histone variant levels and leads to reduced nucleosome occupancy at specific loci. Simultaneous mutation of the H3-H4 chaperone HIRA (Histone Regulator A) and ATRX is detrimental for plant survival, suggesting that HIRA and ATRX function in complementary pathways of H3.3 deposition. Finally, we show that ATRX controls nucleosome occupancy at 45S ribosomal DNA (rDNA) loci, affects nuclear localization of 45S rDNA loci, as well as abundance and expression of 45S rDNA variants. Altogether, our results emphasize the complexity of the Arabidopsis chaperone landscape with ATRX as a novel player and its particular role in the regulation of 45S rDNA expression.

Introduction

Gene regulation of the eukaryotic genome necessitates a controlled balance between packaging of the large linear DNA molecules and permitting regulated access to protein complexes involved in DNA transcription, replication and repair. Within nucleosomes, the basic building blocks of chromatin, the DNA double-strand helix wraps around octamers of histone proteins. Canonical histones are deposited on newly synthesized DNA to maintain nucleosomal density following passage of the replication fork. These canonical histones can then be replaced with specific histone variants to modify nucleosome composition, stability, higher-order chromatin organization and DNA accessibility in a site-specific manner. Most eukaryotes express variants of the canonical histone H3.1, such as the replacement variant H3.3, as well as tissue-specific H3 variants (Talbert et al., 2012). In mammals, H3.3 deposition is associated with dynamic chromatin regions such as transcriptionally active genes and regulatory regions, high nucleosome turnover and DNA accessibility. The plant model *Arabidopsis thaliana* encodes also H3.1 and H3.3 proteins that differ by only four amino acids, as well as H3.3-like variants expressed in specific reproductive tissues (Ingouff et al., 2010; Okada et al., 2005). Genome-wide studies revealed preferential enrichment of H3.1 at heterochromatic regions and of H3.3 at active genes and telomeric repeats (Stroud et al., 2012; Wollmann et al., 2012; Vaquero-Sedas and Vega-Palas, 2013; Shu et al., 2014).

To ensure the incorporation of the adequate histone type at the right time and genomic location, specialized proteins called histone chaperones associate with histones during their shuttling from cytoplasm to nucleoplasm and deposit histones on DNA. These histone chaperones operate in a coordinated network. For the assembly of H3-H4 dimers into nucleosomes, several chaperone complexes have been characterized in eukaryotes. The CAF-1 (Chromatin Assembly Factor 1) complex deposits H3 in a DNA-synthesis-coupled manner during DNA replication and repair (Tagami et al., 2004; Smith and Stillman, 1989) and is composed of the subunits FAS1 (FASciata1), FAS2 (FASciata2) and MSI1 (Multicopy Suppressor of Ira1) in Arabidopsis (Kaya et al., 2001). The HIR complex incorporates H3 histones in a DNA synthesis-independent manner during the whole cell cycle or in resting cells. In Arabidopsis, this complex is composed of HIRA (HIstone Regulator A), UBN1/UBN2 (UBiNuclein 1/2) and CABIN1 (CAlcineurin BINding protein 1) (Nie et al., 2014; Duc et al., 2015). In mammals, these chaperone complexes show clear variant specificity, with CAF-1 depositing the canonical histone H3.1, while the HIR complex assembles H3.3 (Tagami et al., 2004). Little is known so far concerning such a specificity of the different complexes for H3 histone variants in plants, except that the Arabidopsis HIR complex binds H3.3 (Nie et al., 2014). In mammals, additional proteins that function as histone chaperones have been described, such as the DEK (Sawatsubashi et al., 2010) and the ATRX/DAXX (Alpha Thalassemia-mental Retardation X-linked syndrome / Death-domain Associated protein) hetero-complex which shows specificity for H3.3 (Drané et al., 2010; Goldberg et al., 2010; Lewis et al., 2010; Wong et al., 2010).

ATRX/DAXX and HIR show differential chromatin-binding patterns (He et al., 2015; Pchelintsev et al., 2013) and are known to deposit histone H3.3 at distinct genomic regions. Indeed, while HIRA

deposits H3.3 at genic regions (Pchelintsev et al., 2013; Goldberg et al., 2010; Ray-Gallet et al., 2011), ATRX/DAXX incorporates this histone variant at pericentromeric repeats, telomeres, endogenous retroviral elements and silenced imprinted alleles (Elsässer et al., 2015; Voon et al., 2015; Filipescu et al., 2013; Goldberg et al., 2010; Lewis et al., 2010; Gokhman et al., 2013; Xue et al., 2003; McDowell et al., 1999). ATRX has the capacity to bind lysine 9 (K9) methylated histone tails *via* its ADD (ATRX-DNMT3-DNMT3L) N-terminal domain (Iwase et al., 2012; Eustermann et al., 2011). This function is thought to contribute to the specific deposition pattern of H3.3 histones. Besides the ADD domain, ATRX contains a carboxy-terminal SWI/SNF2-like ATPase motif found in chromatin remodeling proteins. These proteins use the energy of ATP to modulate histone-DNA interactions within nucleosomes and contribute to a wide range of cellular processes including recombination (Alexeev et al., 2003), DNA replication (Collins et al., 2002), and histone exchange (Mizuguchi et al., 2004; Konev et al., 2007). The human ATRX has been described about two decades ago due to various disorders associated with mutations in the *ATRX* gene such as the X-linked-thalassemia mental retardation syndrome, characterized by several development abnormalities (Gibbons et al., 1995b, 1995a). Besides its role as component of a histone chaperone complex, several studies analyzed the effects of ATRX loss of function in mammals and defined additional roles for this protein including telomere maintenance – notably by protecting them from replication fork stalling – DNA replication – defects in ATRX leading to a prolongation of the S phase –, and heterochromatin silencing (reviewed in (Clynes et al., 2013)). Indeed, several recent studies connect DAXX/ATRX with silencing of retrotransposons and satellite sequences via the incorporation of H3.3 and the recruitment of histone methyltransferase activity (Sadic et al., 2015; Voon et al., 2015; He et al., 2015; Elsässer et al., 2012).

The ATRX gene has been conserved through evolution but intriguingly, in invertebrates such as Drosophila, ATRX is split into two proteins, dADD1 (CG8290) which harbors the human ATRX ADD domain (López-Falcón et al., 2014; Alekseyenko et al., 2014) and dATRX/XNP (X-linked Nuclear Protein) which contains the SWI/SNF-like ATPase domain (Bassett et al., 2008; Emelyanov et al., 2010). The dATRX/XNP protein shares similar functions to its mammalian counterpart; however, XNP isoforms localize to chromosome arms and heterochromatin regions (Bassett et al., 2008). In human and mouse, ATRX null alleles are embryo-lethal (Garrick et al., 2006), but semi-lethal for fly (Lee et al., 2007). In plants, despite the identification of a putative ATRX coding gene in Arabidopsis (Shaked et al., 2006; Otero et al., 2014), the involvement of this protein in histone H3 deposition and chromatin function has not been analyzed yet.

In this study, we investigated the potential role of ATRX in histone H3 deposition and chromatin function in *Arabidopsis thaliana*. We showed that the Arabidopsis *atrx* null mutants are viable despite developmental defects and display moderate cell cycle defects. Genetic analyses using a combination of mutants in the different histone H3 incorporation systems suggest that ATRX functions as a histone H3 chaperone. Indeed, loss of ATRX affects cellular histone H3 pools and H3 incorporation at certain genomic loci. Most importantly, we provide evidence that ATRX controls nucleosome occupancy at 45S

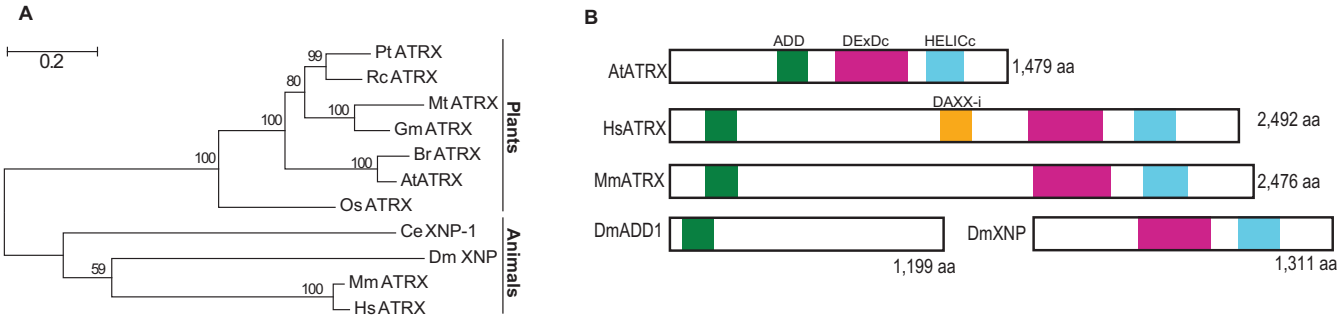


Figure 1. Identification of an ATRX ortholog in *Arabidopsis*.
(A) Phylogenetic tree of ATRX proteins. Pt, *Populus trichocarpa*; Rc, *Ricinus communis*; Mt, *Medicago truncatula*; Gm, *Glycine max*; At, *Arabidopsis thaliana*; Br, *Brassica rapa*; Os, *Oryza sativa*; Mm, *Mus musculus*; Hs, *Homo sapiens*; Dm, *Drosophila melanogaster*; Ce, *Caenorhabditis elegans*. Scale bar is 0.2 substitutions/per site. The percentage of trees in which the associated taxa clustered together is shown next to the branches.
(B) Functional domains of ATRX proteins. ADD, ATRX-DNMT3-DNMT3L; DAXX-I, DAXX-interacting domain; DEXDc, DEAD-like helicase superfamily; HELICc, HELICase superfamily C-terminal domain.

rDNA loci, 45S rRNA gene transcription and variant dosage.

Results

Identification and characterization of the Arabidopsis ATRX ortholog

Using the conserved domains of the human ATRX, we confirmed the identification of At1g08600, also called CHR20 (Chromatin Remodelling 20) (Shaked et al., 2006; Otero et al., 2014), as the unique Arabidopsis AtATRX ortholog. Plant ATRX proteins share similarities with mammalian and invertebrate ATRX proteins although they form a distinct monophyletic group (Figure 1A). Interestingly, as in mammals, the Arabidopsis ATRX harbors simultaneously the N-terminal ADD domain and the C-terminal helicase domain, contrary to invertebrates, such as Drosophila, that express two distinct proteins, dADD1 (López-Falcón et al., 2014; Alekseyenko et al., 2014) and dATRX/XNP (Figure 1B). The plant protein is much shorter than its mammalian counterpart and devoid of the large central region involved in DAXX interaction (Tang et al., 2004) (Figure 1B). Based on the built 3D model (Supplemental Figure S1D), we emphasize that the plant and the human ADD domains both contain the GATA and PHD zinc finger helices and share 36.8% similarity for this domain (Supplemental Figure S1A&D). The C-terminal region of AtATRX contains the DExDc and the HELICc subdomains constituting the ATPase domain required for chromatin remodeling and characteristic of the SNF2 family proteins (46.2% and 54.8% similarity, respectively, for DExDc and HELICc domains with HsATRX, Supplemental Figure S1B-C). Taken together, the Arabidopsis ATRX protein shares major features with its human counterpart, leading us to explore its potential role in histone dynamics and assembly.

We first quantified transcript levels of *ATRX* by qRT-PCR, which showed expression in all tested plant tissues (Supplemental Figure S1E). From the stock collection, we ordered two different mutant alleles in which T-DNAs are inserted in an exon (Figure 2A) and established homozygous mutant plants. RT-PCR and qRT-PCR analysis revealed that the *atrx-1* allele is a knockout mutant (Figure 2B-C and Supplemental Figure S2B). The remaining transcript would encode only a short truncated protein devoid of the ADD and helicase domains (Supplemental Figure S2C). The *atrx-2* allele produces remaining transcripts including part of the *ATRX* 5' and 3' region (Figure 2B and Supplemental Figure S2B-C), suggesting that truncated versions of the ATRX protein containing the majority of the ADD or part of the HELICc domain could be produced.

Both mutants are viable and show no obvious vegetative abnormality compared to wild type (WT) plants (Figure 2D), except a reduced rosette surface for the *atrx-1* allele (Figure 2E) and reduced root growth in both mutant alleles (Fig 2F). In addition, siliques of *atrx-1* and *atrx-2* have reduced viable seed content and increased amount of aborted seeds, and more unfertilized ovules were scored in the *atrx-2* mutant plants (Figure 2G-H). The *atrx* anthers develop normally but Alexander staining revealed some non-viable pollen grains in the *atrx-2* allele (Figure 2I) and reduced pollen content in both *atrx-1* and *atrx-2* mutant plants. Given that in mammals, ATRX mutations are associated with replication defects (Leung et al., 2013; Huh et al., 2012), we performed FACS analysis on dissected whole

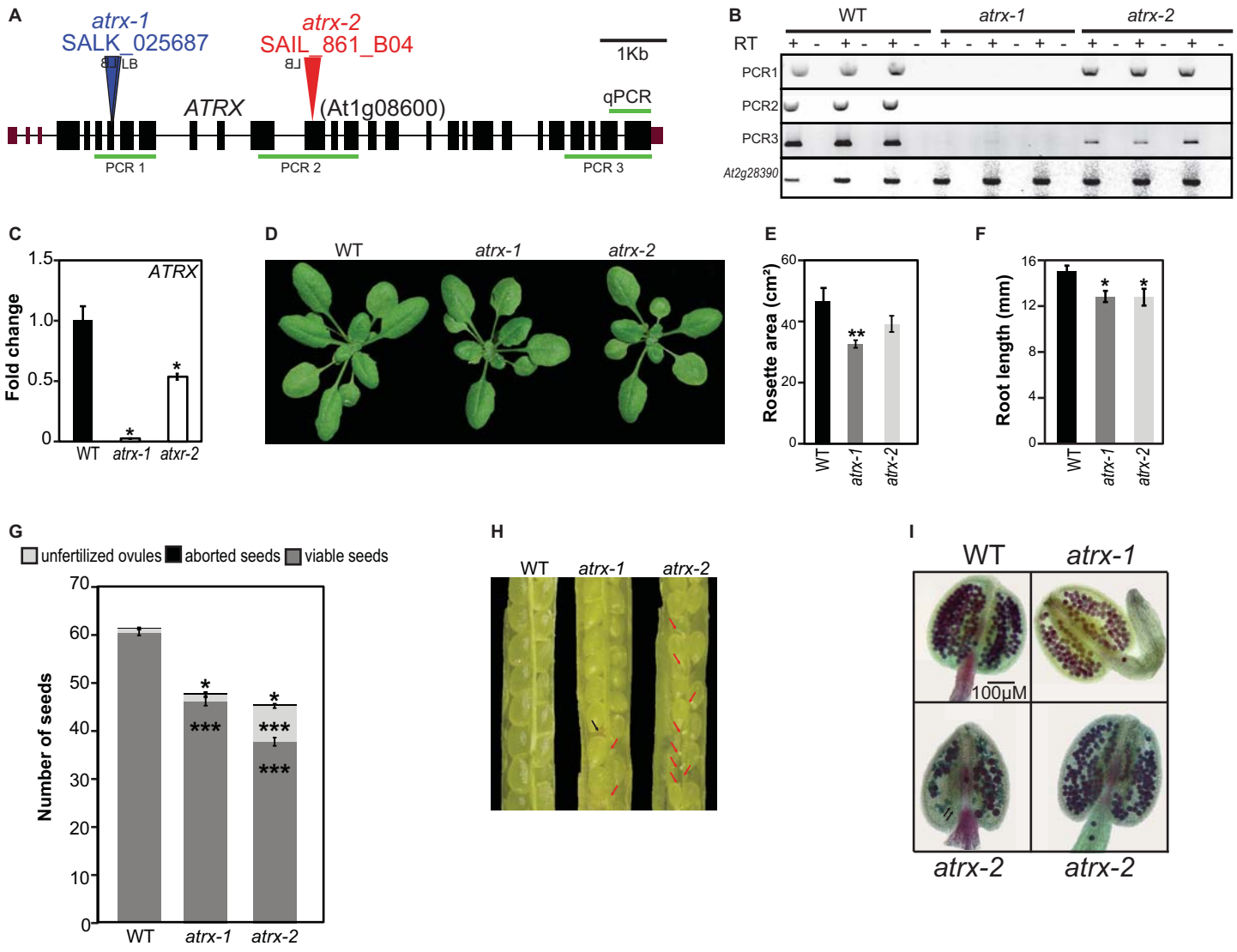


Figure 2. Characterization of the Arabidopsis atrx mutant alleles.

(A) Gene structure of the Arabidopsis ATRX. Exons, black rectangles; UTRs, purple rectangles; introns, lines; T-DNA insertion, triangle; LB, left border.

(B) Analysis of ATRX transcripts produced in mutants with the atrx-1 or atrx-2 alleles by semi-quantitative RT-PCR on three biological replicates. The amplified regions are displayed by green lines in (A). At2g28390 was used as a control.

(C) Expression of ATRX in atrx mutants analyzed by qRT-PCR. The analyzed region is displayed in (A).

(D) Representative 3-week-old atrx mutant plantlets grown on soil.

(E) Quantification of rosette surface area of atrx mutant plants. Quantification was done on at least six 2-week-old plants for each genotype. Student test, **, $p<0.01$.

(F) Root length quantification of atrx mutant plants. Quantification was done on at least 4 5-day old in vitro grown plants for each genotype. Student test, *, $p<0.05$.

(G) Quantification of seed content in atrx mutant siliques compared to WT. Quantification is based on at least 30 siliques pooled from 4 plants. Student test, *, $p<0.05$; ***, $p<0.001$.

(H) Representative dissected siliques from atrx mutants. Red arrows indicate unfertilized ovules and the black arrow an aborted seed.

(I) Pollen viability assessed by Alexander staining. Both atrx-1 and atrx-2 anthers show reduced pollen content, and atrx-2 anthers contain non-viable pollen (green color) indicated by black arrows.

cotyledons. For the *atrx-1* allele, we noticed increased endoreplication (Supplemental Figure S2D) and broader peaks in the FACS profile (Supplemental Figure S2E) that might be caused by delayed S-phase progression or replication defects. While we found no hypersensitivity to hydroxyurea (HU) (Supplemental Figure S2F), we noted an increased number of nuclei in early compared to late S-phase (Supplemental Figure S2H) after EdU (5-ethynyl-2'-deoxyuridine) staining of root apexes (Supplemental Figure S2G). This might be indicative of slower progression through early S-phase. Furthermore, a previous study has implicated ATRX in the DNA damage response using an RNAi line (Shaked et al., 2006). We confirmed moderate sensitivity to gamma-irradiation in our *atrx* mutant alleles using a DNA-damage sensitivity assay, which monitors emergence of true leaves after seed irradiation (Supplemental Figure S2H).

Taken together, *atrx* mutant alleles are viable but display moderate growth, reproductive and replication defects.

Epistatic relationships between players in the H3 incorporation pathways

Since ATRX is an essential component of a histone chaperone complex in several organisms and is highly conserved through evolution, we hypothesized that ATRX may play a similar role in Arabidopsis. Hence, we investigated the relationships between known Arabidopsis histone H3 chaperone complexes and ATRX for epistasis, by crossing mutants of the HIR complex (consisting of UBN2, CABIN and HIRA) (Nie et al., 2014; Duc et al., 2015) and the CAF-1 complex (Kaya et al., 2001) to each *atrx* mutant.

For the *hira-1* mutants that lack the central subunit of the complex (Duc et al., 2015), we obtained in the F2 generation only one double mutant plant from the *atrx-1* cross (n=111) and none from *atrx-2* cross (n=115, Supplemental Table S1A-B), although the respective loci are unlinked. We confirmed the distorted segregation by analyzing the F3 progeny (Supplemental Table S2A-B). While *hira-1* and *atrx-1* single mutants display no vegetative phenotype, the few *hira-1 atrx-1* double mutants recovered are sterile and severely affected in growth (Figure 3A). Furthermore, different combinations of F2 plants heterozygous for one mutation and homozygous for the other show normal flowers (Supplemental Figure S3A-B), but reduced number of pollen grains (Supplemental Figure S3C-D) and exacerbated seed set defects were observed in *hira-1 atrx-2/ ATRX* compared to *hira-1* (Figure 3B and Supplemental Figure S3E). Together, this shows that the simultaneous loss of HIRA and ATRX impairs plant viability or causes severe developmental defects in the surviving plants. For the crosses with mutants for the other two subunits of the HIR complex, UBN2 and CABIN1, we obtained viable double mutants with expected or reduced frequency, respectively (Supplemental Tables S1C-F and S2C-D), and these double mutants display a reduced leaf surface (Figure 3C-E and Supplemental Figure S3F-H).

To explore the relationship with the CAF-1 complex, which is thought to be involved in replication-coupled histone deposition, we crossed the two *atrx* mutants with the *fas2-5* mutant, a knock-out mutant for the second largest subunit of the CAF-1 complex (Duc et al., 2015). Double mutants were obtained with the expected frequencies (Supplemental Table S3). When we looked closer at the development of the double

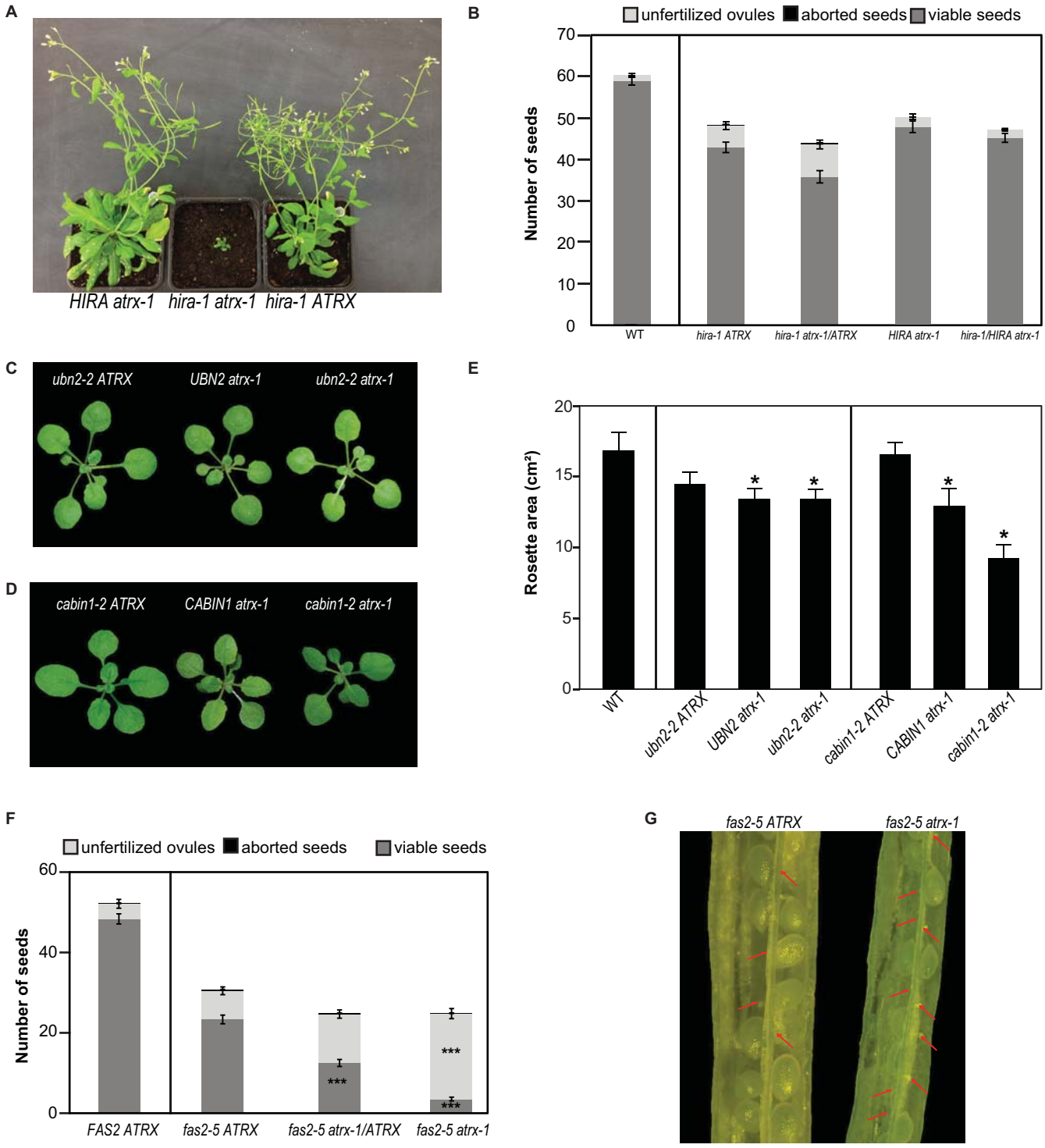


Figure 3. Epistatic relationship between ATRX and CAF-1 or HIR histone chaperone complexes.
(A) Representative 7-week-old F3 sister plants grown on soil.
(B) Quantification of seed content in hira-1, hira-1 atrx-1/ATRX, atrx-1 and atrx-1 hira-1/HIRA mutant siliques. Quantification is based on at least 30 siliques pooled from 4 plants.
(C-D) Representative 3-week-old F3 plants derived from crosses of atrx-1 alleles with ubn2-2 (C) or cabin1-2 (D) grown on soil.
(E) Quantification of total rosette surface area of the F3 sister plants described in (C) and (D). Quantification is based on at least six 2-week old plants for each genotype. Student test, *, $p<0.05$.
(F) Quantification of seed content in siliques of FAS2 ATRX, fas2-5 ATRX, fas2-5 atrx-1/ATRX and fas2-5 atrx-1 F2 sister plants. Quantification is based on at least 30 siliques pooled from 4 plants. Student test in comparison to fas2-5 ATRX, ***, $p<0.001$.
(G) Representative dissected siliques from fas2-5 ATRX and fas2-5 atrx-1 F2 sister plants. Red arrows indicate unfertilized ovules.

mutant plants, we noticed flowers similar to *fas2-5* (Supplemental Figure S3I-J), but with aggravated anther shapes (Supplemental Figure S3K-L), and further reduction of the already low seed set (Figure 3F-G and Supplemental Figure S3M-N).

Taken together, the analysis of the crosses between mutants of the known H3 chaperone complexes CAF-1 and HIR with *atrx* mutant alleles reveals aggravated developmental defects and severe growth deficiencies or lethality, respectively. In particular, the lethality observed in combination with mutants of the replication-independent system of H3 incorporation mediated by the HIR complex suggests that Arabidopsis ATRX could play a complementary role in variant histone interaction and deposition in plants.

The Arabidopsis ATRX displays features of a histone chaperone

Based on the described results, we reasoned that loss of Arabidopsis ATRX function might affect histone H3 pools. We used the knockout allele *atrx-1* to investigate by western blots how different H3 histones pools are affected in comparison to *hira-1*. We used two different protocols (Durut et al., 2014; Honda et al., 1966) to recover histones from distinct cellular fractions (see Material and Methods for details). In comparison to WT plants, non-nucleosomal H3 amounts are reduced in *atrx-1* (Figure 4A) to similar levels as previously shown for *hira-1* (Duc et al., 2015), and this reduction occurred without changes in the expression levels of several tested H3.1 and H3.3 histone genes in the *atrx-1* mutant (Supplemental Figure S4A). Reduced H3 levels were also observed in total extracts and nuclear fractions (Figure 4B). This result suggests that loss of either HIRA or ATRX impacts histone pools.

We then further investigated the possible histone chaperone function of ATRX by analyzing nucleosome occupancy at specific genomic sites in WT, *atrx-1* and *hira-1* mutants. For this purpose, we combined H3-ChIP with quantitative PCR on 2.5-week-old *in vitro*-grown plants. We first looked at heterochromatic targets and chose two centromeric regions: the 180bp repeats and the 106B long terminal (LTR)-like repeats (Thompson et al., 1996; Fransz et al., 1998), and two pericentromeric regions: TSI (Transcriptionally Silent Information) (Steimer et al., 2000) and the 5S ribosomal DNA loci (Benoit et al., 2013; Cloix et al., 2000; Campell et al., 1992). In comparison to *hira-1* that leads to reduction in nucleosome occupancy at TSI and 106B (Duc et al., 2015), nucleosomal occupancy for these four analyzed heterochromatic regions was maintained in *atrx-1* (Figure 4C). We then analyzed loci on chromosome arms: the Mutator-like DNA transposon (*Mule*, At2g15810) and three transcriptionally active genes (*UBC28*, *UEV1C* and *HXK1*) with different expression levels (Duc et al., 2015) and found that H3 incorporation was decreased at *HXK1* (Figure 4D). Finally, we analyzed three subtelomeric regions, given that ATRX is enriched at telomeres and subtelomeric regions of human chromosomes (Law et al., 2010). We observed moderately reduced H3 occupancy only at At5g67640, a gene in the subtelomeric region of the long arm of chromosome 5 (Figure 4E).

In mammals, ATRX is involved in H3.3 deposition at pericentromeric regions, which triggers H3K9 methylation (He et al., 2015; De La Fuente et al., 2015). This prompted us to analyze the H3K9me2 mark,

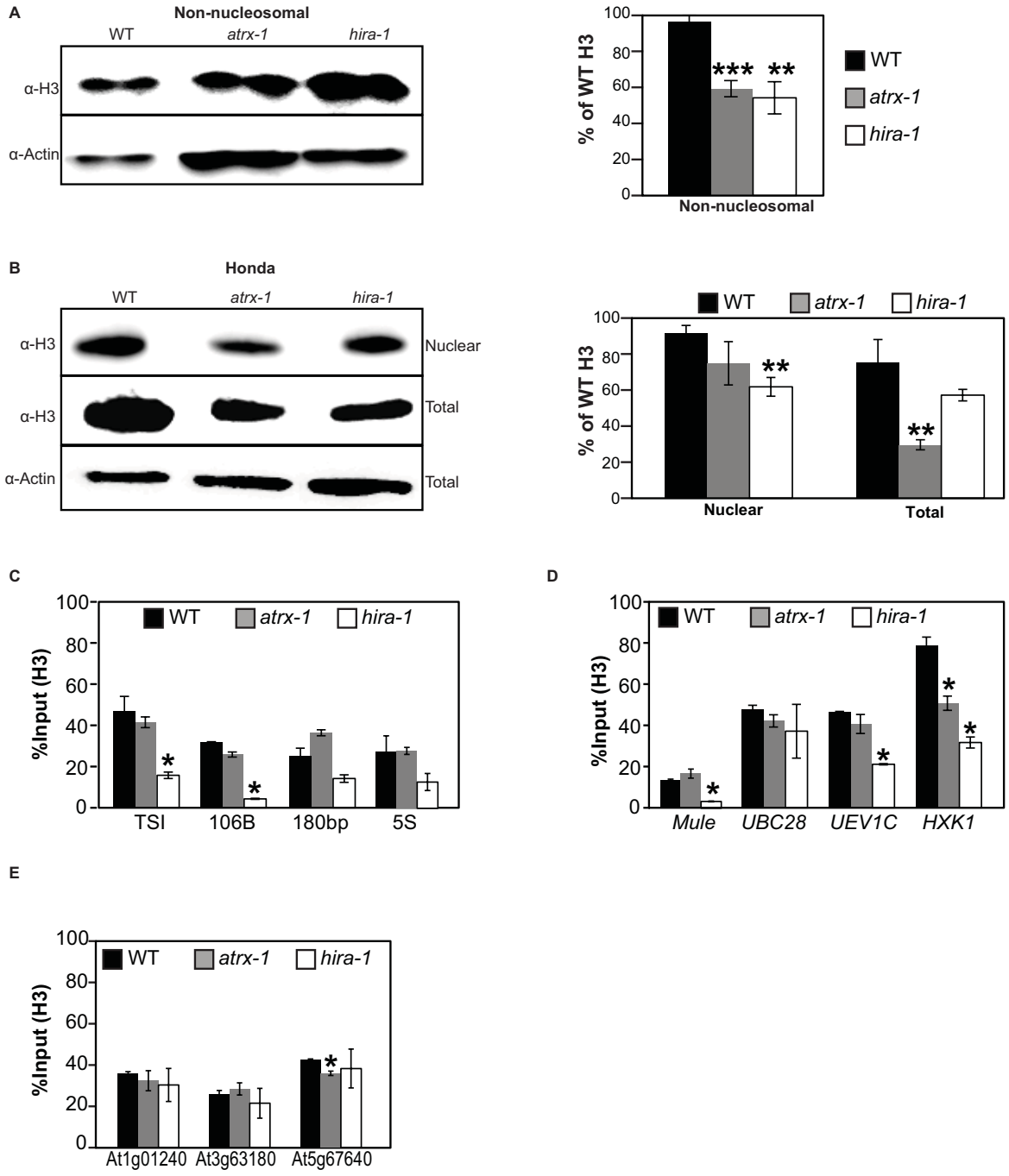


Figure 4. Effects of ATRX loss on histone pools and nucleosome occupancy.

(A) Left: Histone H3 protein levels quantified by western blot in non-nucleosomal fractions. Twenty micrograms of proteins extracted from 2.5-week old in vitro-grown plants were loaded per lane. Right: Quantification of H3 band intensities normalized to actin from 2 independent experiments and several blots. Student test, **, p<0.01; ***, p<0.001.

(B) Left: Histone H3 protein levels quantified by western blot in nuclear fraction and total extracts. Six micrograms of proteins extracted from 2.5-week-old in vitro-grown plants were loaded per lane. Right: Quantification of H3 band intensities normalized to actin in total extracts from 2 independent experiments and several blots. Student test, **, p<0.01.

(C-D-E) Histone H3 occupancy at centromeric and pericentromeric repeats (TSI, 106B, 180bp, ribosomal 5S rDNA loci) (C), at one transposon on a chromosome arm (At2g15810, Mule), at three active genes (UBC28, UEV1C, HXK1) (D) and at three subtelomeric regions (At1g01240, At3g63180, At5g67640) (E) assessed by H3-ChIP qPCR in 2.5-week-old in vitro-grown WT, atrx-1 and hira-1 mutant plants. Student test, *, p<0.05.

the major plant heterochromatin signature present in pericentromeric regions and in patches of heterochromatin on chromosome arms (Bernatavichute et al., 2008). No difference in H3K9me2 levels relatively to H3 was observed at the heterochromatin repeats TSI, 106B and 180bp or at Mule (a heterochromatin locus on a chromosome arm deregulated in *hira-1* mutants (Duc et al., 2015)) and at the telomeric region with decreased H3 occupancy (At5g67640, Supplemental Figure S4B). This suggests that ATRX does not impact the setting or maintenance of this repressive histone mark at these sites. In agreement with this observation, silencing of the centromeric and pericentromeric 180bp, 106B and TSI repeat sequences (Supplemental Figure S4C) or of different transposable elements (list of tested loci available in Material and Methods) was maintained, except for the *Mule* transposon that was reactivated in *atrx-1* (Supplemental Figure S4D).

Therefore, loss of ATRX affects histone pools and has a moderate effect on nucleosome occupancy at certain active genes and subtelomeric regions. In contrast, it is neither required for nucleosome occupancy at the tested pericentromeric regions nor to maintain transcriptional silencing at the analyzed repeats and at most transposons tested.

Interplay between Arabidopsis ATRX and H3.1 histones

In mammals, the ATRX/DAXX complex shows specificity for the H3.3 variant (Elsässer et al., 2012; Liu et al., 2012). We therefore wanted to analyze whether loss of ATRX in Arabidopsis specifically affects H3.1 or H3.3 pools and their chromatin distribution. Arabidopsis canonical histones H3.1 and variants H3.3 differ by only 4 amino acids, and no variant-specific antibodies are available so far in plants. Hence, we followed the incorporation of both H3 histone types by expressing FLAG-HA-tagged versions of H3.1 (epitope-tagged H3.1, eH3.1) or H3.3 (eH3.3) under the control of their endogenous promoters. We first validated the incorporation of eH3.1 and eH3.3 proteins by FLAG-ChIP. As expected, eH3.1 is preferentially enriched at heterochromatic loci (180bp, TSI and the *Ta3* retrotransposon localized in the pericentromeric region of chromosome 1), while transcriptionally active genes are enriched in eH3.3 (*UEV1C*, *UBC28*, *HXK1*) (Figure 5A). We then crossed the eH3.1 and eH3.3 transgenic lines with the *atrx-1* mutant. From the progeny, we selected sister plants with and without the *atrx-1* mutation. Unfortunately, in the progeny from two independent F2 lines, we were unable to identify plants that retained eH3.3 expression, so that we could not study the effect of ATRX disruption on eH3.3 incorporation. In the eH3.1 progeny, we succeeded to select WT and mutant sister plants that expressed sufficient eH3.1 for further analysis. We chose two lines for both genotypes with similar expression levels of the transgene (Supplemental Figure S5A), to ensure that changes in eH3.1 abundance, further investigated by ChIP and western blots, are caused by ATRX loss and not by different transgene expression levels.

We first investigated by western blots on 2.5-week-old soil-grown plants how eH3.1 histones pools are affected by the absence of ATRX. Since the anti-H3 (□-H3) antibody recognizes endogenous H3.1 and H3.3 as well as eH3.1, blots were hybridized with a □-H4 antibody and the ratio of eH3.1 relative to H4 was determined. With the □-H4 antibody, we also confirmed the reduced H3-H4 pool in the total protein extract in the *atrx-*

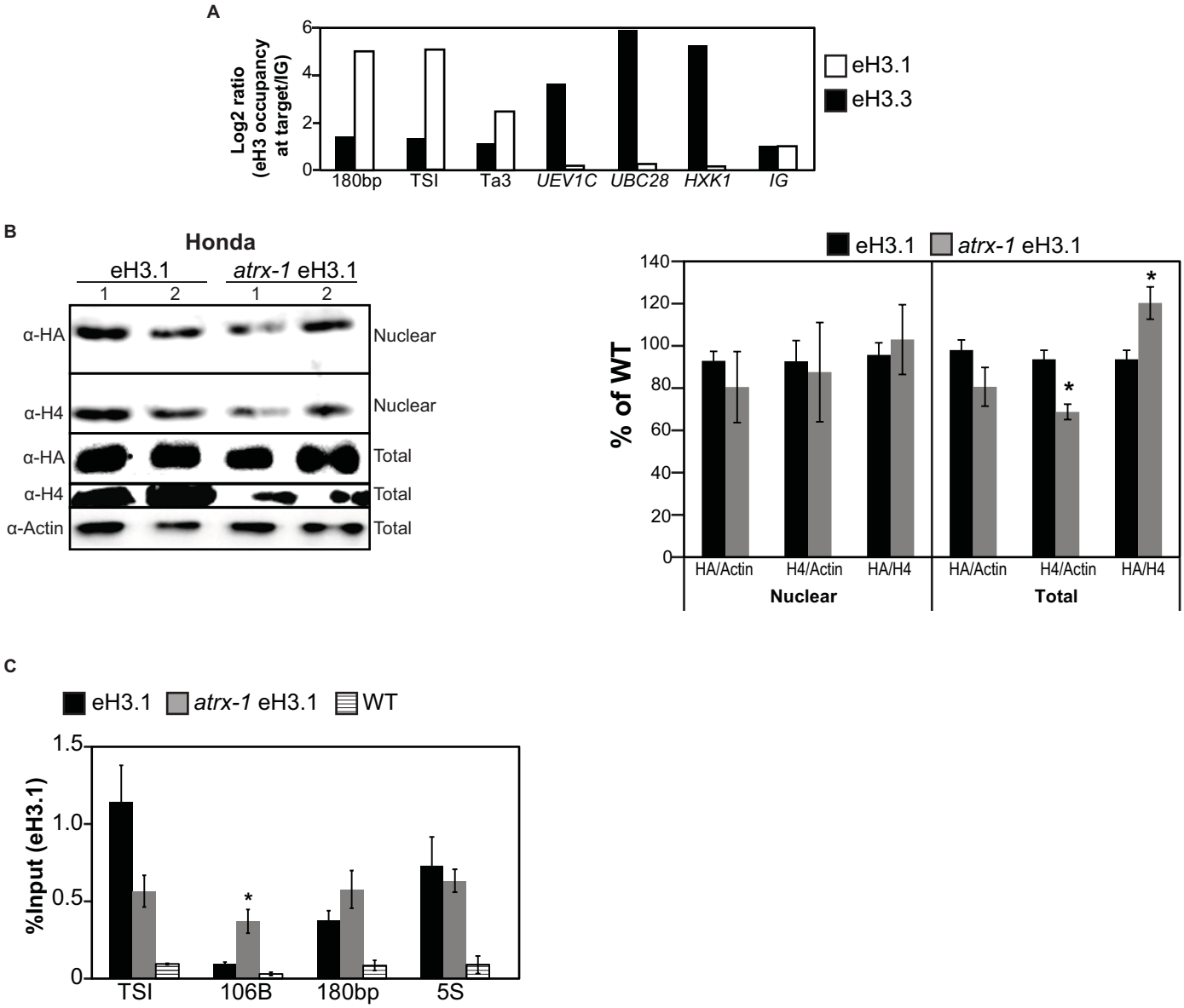


Figure 5. Impact of ATRX loss on incorporation and balance of histone H3 variants.
(A) Ratio of canonical eH3.1 and variant eH3.3 occupancy at heterochromatic repeats (180bp, TSI, Ta3) and at three active genes (UBC28, UEV1C, HXK1) assessed by FLAG-ChIP qPCR. The eH3 occupancy is normalized to the intergenic region (IG, set to 1).
(B) Left: Histone H3 protein levels quantified by western blot in nuclear fraction and total extracts. Six micrograms of proteins extracted from 2.5-week-old in vitro-grown plants were loaded per lane. Right: Quantification of HA and H4 band intensities relative to actin quantified in total extracts from 2 independent experiments and several blots. Student test, *, $p<0.05$.
(C) Histone H3.1 occupancy at heterochromatic repeats (TSI, 106B, and 180bp) and at the 5S ribosomal DNA loci assessed by FLAG-ChIP qPCR in 2.5-week-old in vitro-grown eH3.1, atrx-1 eH3.1 and WT plants. Student test, *, $p<0.05$.

1 eH3.1 lines (Figure 5B). Furthermore, we found an increased amount of eH3.1 relatively to H4 in the total extract of *atrx-1* eH3.1 (Figure 5B). We then performed ChIP coupled to quantitative PCR on 2.5-week-old *in vitro*-grown eH3.1 and *atrx-1* eH3.1 plants. We first validated these lines by performing H3-ChIP and observed similar H3 enrichment patterns for *atrx-1* eH3.1 as in *atrx-1* (Supplemental Figure S5C). Then, in order to determine potential changes in the incorporation of eH3.1, we performed FLAG-ChIP using WT plants as a control for background noise (Figure 5C). We evaluated eH3.1 levels at heterochromatic regions TSI, 106B, 180bp known to be enriched in H3.1 (Duc et al., 2015) and at 5S rDNA. In *atrx-1* mutants, the eH3.1 abundance is maintained at TSI, 180bp and 5S rDNA regions, while 106B repeats display a higher abundance of eH3.1 (Figure 5C). Together, these data show that H3.1 incorporation is maintained in the absence of ATRX suggesting that altered H3.3 handling is the cause for the reduced total H3 amounts.

The Arabidopsis ATRX maintains H3 occupancy at 45S rDNA genes and modulates their variant dosage

During the ChIP analysis of *atrx-1* mutants, we further noticed altered nucleosomal occupancy at the tandem arrays of 45S rDNA loci (Figure 6A). In Arabidopsis, each ribosomal 45S gene transcription unit encodes a precursor transcript that includes the structural rRNAs (18S, 5.8S, 25S) and the Internal and the External Transcribed Spacers (ITS and ETS). Our H3-ChIP combined with quantitative PCR showed reduced H3 levels both at the 25S and 18S regions (Figure 6A). Since heterochromatin signatures have been reported to be altered at sites affected by ATRX disruption (Voon et al., 2015; Sadic et al., 2015), we also analyzed the H3K9me2 mark, the major plant heterochromatin signature, at 45S rDNA loci. To our surprise, we noticed a general increase in H3K9me2 enrichment relatively to H3 in the *atrx-1* mutant at 45S loci (Figure 6B). Three major 45S sequence variants (VAR1 to 3) defined by small sequence variations in the ETS region have been described in Arabidopsis, a fourth one is expressed but not distinguishable when amplified from genomic DNA (Pontvianne et al., 2010; Durut et al., 2014). Out of the 4 variants, mainly VAR2 and VAR3 are generally expressed in WT plants. Given the reduced nucleosome occupancy at 45S rDNA and the enrichment in H3K9me2, we wanted to know whether the balance in the expression of 45S rDNA variants is affected in *atrx* mutants. Indeed, we noticed that the expression profile of the different variants is changed in both *atrx-1* and *atrx-2* mutant alleles compared to WT, with the VAR2 being predominantly expressed (Figure 6C, Supplemental Figure S6A). To explain this observation, we looked at the relative abundance of the rDNA variants, as well as their enrichment in H3 or H3K9me2, by submitting ChIP samples to PCR (Figure 6D-E). Using a primer pair that amplifies all variant types (Pontvianne et al., 2010), the different variants can then be separated by size and their relative abundance quantified. In WT, VAR1 is enriched in H3 and H3K9me2 when compared to the ChIP input DNA, while VAR3 is relatively depleted (Figure 6E). This is in agreement with previous observations showing that VAR1 is in general transcriptionally silent, but reactivated in mutants affecting H3K9me2 levels (Pontvianne et al., 2012). This pattern can also be observed in *atrx-1.* When comparing WT and *atrx-1* mutants, we noticed that VAR2 is more abundant in the *atrx-1* genome than in WT plants, while VAR3 is

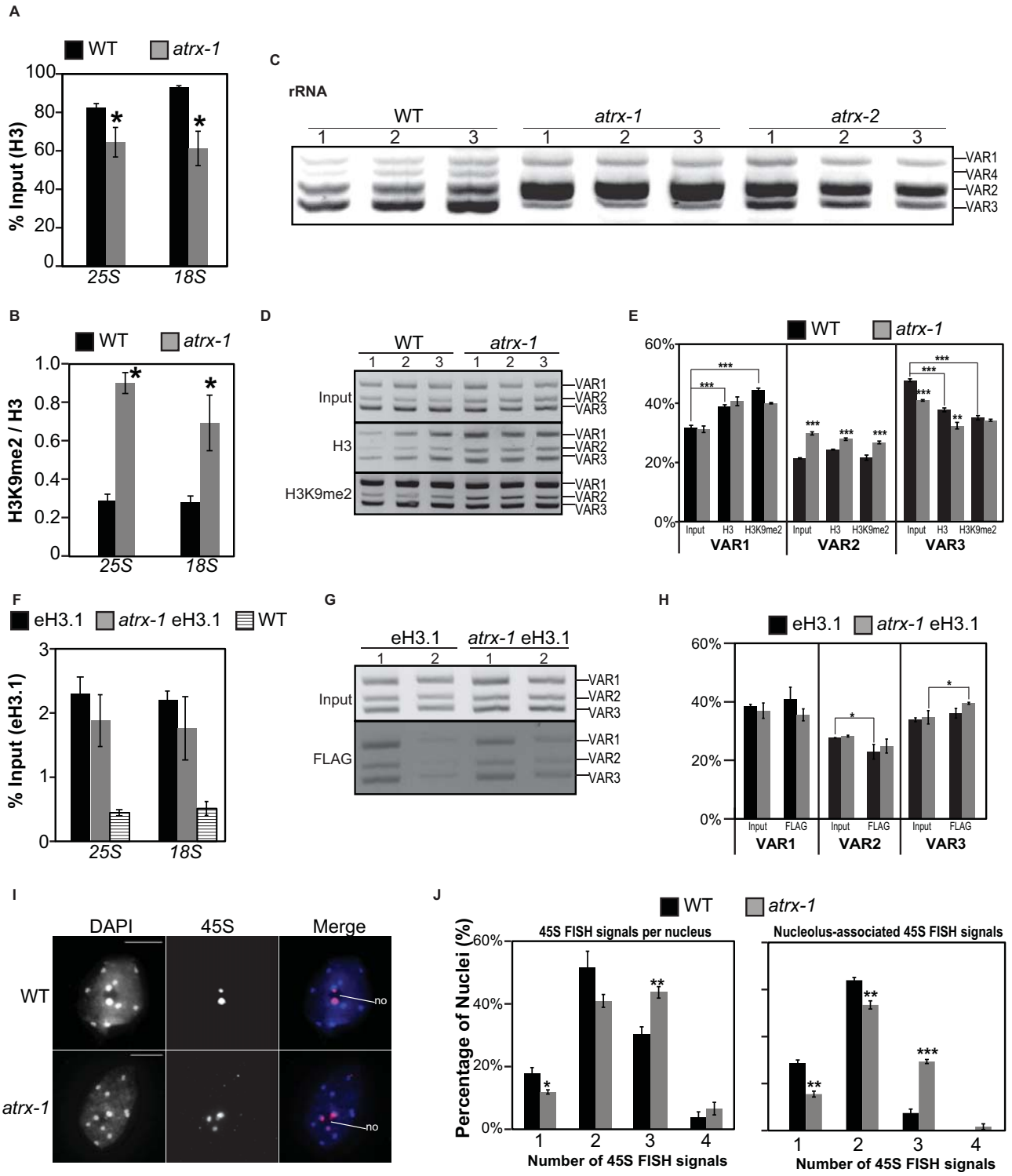


Figure 6. Effects of ATRX loss on nucleosome occupancy and balance between histone variants at 45S loci.

(A) Histone H3 occupancy at 25S and 18S rDNA loci assessed by H3-ChIP qPCR in 2.5-week-old *in vitro*-grown WT and *atrx-1* mutant plants. Student test, *, p<0.05.

(B) Ratio of histone H3K9me2/H3 occupancy at 25S and 18S rDNA loci assessed by ChIP qPCR in 2.5-week-old *in vitro*-grown WT, *atrx-1* and *hira-1* mutant plants. Student test, *, p<0.05.

(C) RT-PCR on cDNA prepared from three biological replicates of 2.5-week-old i*n vitro*-grown WT, *atrx-1* and *atrx-2* plants. Detection of pre-rRNA transcripts was performed with primers enclosing the external transcribed 45S spacer and detecting all variants.

(D) Analysis of 45S variant abundance in the input, H3-ChIP and H3K9me2-ChIP performed on three biological replicates from 2.5-week-old *in vitro*-grown WT and *atrx-1* mutant plants.

(E) Quantification of relative band intensities presented in (D) and amplified from the input, H3-ChIP and H3K9me2-ChIP. Student test, **, p<0.01; ***, p<0.001.

(F) Histone H3.1 occupancy at 25S and 18S rDNA loci assessed by FLAG-ChIP qPCR in 2.5-week-old *in vitro-grown* eH3.1, *atrx-1* eH3.1 and WT plants.

(G) Analysis of 45S variant abundance in the input and FLAG-ChIP performed on three biological replicates from 2.5-week old *in vitro* grown eH3.1 and *atrx-1* eH3.1 plants.

(H) Quantification of relative band intensities presented in (G) and amplified from the input and FLAG-ChIP. Student test, *, p<0.05.

(I) FISH for 45S rDNA (purple in merge) in WT and *atrx-1* mutant nuclei. DNA is counterstained with DAPI (in blue in merge). Nucleolus, no.

(J) Quantification of 45S rDNA signals per nucleus (left panel) and of 45S rDNA signals associated with the nucleolus (right panel) in WT and *atrx-1* mutant cotyledon nuclei. At least 350 nuclei obtained from three independent experiments were scored. Student test, **, p<0.01; ***, p<0.001.

underrepresented (Figure 6E), illustrating a change in the relative variant abundance in *atrx-1* mutants. Furthermore, the H3/input and the H3K9me2/input ratios revealed a reduced nucleosome occupancy and H3K9me2 abundance at VAR2 in *atrx-1* (Supplemental Figure S6B). Together, this is consistent with a higher expression of VAR2 in the *atrx-1* mutant.

We then analyzed by FLAG-ChIP the eH3.1 enrichment at 45S rDNA in eH3.1 and *atrx-1* eH3.1 lines using the 25S- and the 18S-specific amplicons. In the *atrx-1* eH3.1 line, we confirmed that H3 occupancy at 45S rDNA is reduced compared to the eH3.1 line (Supplemental Figure S6C). However, no significant change in eH3.1 enrichment was observed on 45S rDNA globally (Figure 6F). When looking at the relative abundance of variants in these lines, we did not observe increased VAR2 abundance compared to WT (Figure 6G-H), likely because these lines are issued from a cross between an eH3.1 expressing plant and an *atrx-1* mutant plant that led to homogenization. Nevertheless, VAR2 is predominantly expressed in *atrx-1* eH3.1 compared to eH3.1 plants (Supplemental Figure S6D-E), confirming the role of ATRX in the regulation of rDNA expression. Furthermore, we found that, compared to input, VAR2 is depleted and VAR3 enriched in eH3.1, in WT and *atrx-1* plants, respectively (Figure 6H), which might contribute to a transcriptional repression of the VAR3 in *atrx-1* mutants.

The 45S rRNA genes are organized in tandem repeats at specific subtelomeric loci situated on chromosome 2 and 4, which form the nucleolus organizer regions (NORs). In interphase, these regions are highly condensed and are part of the chromocenters. Furthermore, NORs coalesce during interphase, and in the majority of cells, two to three NORs can be observed (Durut et al., 2014; Pontvianne et al., 2013, 2012, 2007) (Figure 6I). The NORs that localize close to the nucleolus contain the actively transcribed ribosomal genes that emanate from the highly condensed rDNA loci into the nucleolus. Since we observed changes in 45S variant expression, we determined the number of 45S rDNA FISH signals in WT and *atrx-1* mutants and identified more nuclei with 3 NOR signals compared to WT (Figure 6J, left panel). In addition, the number of NORs in close proximity with the nucleolus increases significantly in *atrx-1* mutants (Figure 6J, right panel), suggesting that the changes in rDNA expression are reflected by a differential organization of NORs in the nucleus. To summarize, ATRX loss affects nucleosome occupancy, H3K9me2 enrichment and nuclear organization of 45S loci and leads to a modified variant dosage and altered 45S rRNA expression levels.

Discussion

We investigated the role of ATRX in the model plant *Arabidopsis thaliana*, using two mutant alleles in the single ATRX gene generated by T-DNA insertions. For both mutant alleles, and in contrast to loss of the mammalian ATRX that causes embryonic lethality, we found that Arabidopsis *atrx* mutants are viable. The *atrx* plants show reduced vigor, slower emergence of true leaves, smaller rosette area and reduced fertility. Both *atrx* alleles result in lethality in combination with HIR complex mutants and affect rDNA expression. Since the *atrx-2* mutants potentially still express the N-terminal part containing nearly the complete ADD domain, the knockout allele *atrx-1* has been chosen for in depth molecular analysis.

ATRX is a multifunctional protein carrying both an ADD domain involved in histone tail binding and a SWI/SNF helicase domain implicated in chromatin remodeling. Therefore, ATRX has a broader role in chromatin structure not necessarily linked to its potential histone chaperone function. Indeed, the moderate sensitivity to gamma-irradiation observed in the mutant plants might be due to its chromatin remodeling activity (Shaked et al., 2006). One of the phenotypes we noted in *atrx-1* mutants is an aberrant cell cycle profile with enlarged peaks in the 2C and 4C populations. An explanation for such a profile might be an S-phase defect. While we found no hypersensitivity to HU, which inhibits replication by diminishing the available nucleotide pool, we noticed more early S-phase nuclei relative to late ones, which could suggest a delayed progression through early S-phase in *atrx-1* mutants. Interestingly, mutants in the CAF-1 chaperone have been suggested to affect S-phase progression by delaying passage through late S-phase (Schönrock et al., 2006). Further studies are needed to exploit fully the role of ATRX in the replication process in plants. However, the binding of mammalian ATRX to genic and intergenic sites rich in variable number tandem repeats that can form G-quadruplexes (Law et al., 2010) may imply that ATRX is needed for the resolution of such structures interfering with proper DNA replication. We also noted increased endoreplication in *atrx* cotyledons, a premature switch to endoreplication occurring also in mutants of the CAF-1 complex that show replication defects (Ramirez-Parra and Gutierrez, 2007) and in response to DNA damage (Adachi et al., 2011).

Lethality and developmental defects observed in combination with mutants for HIR or CAF-1 histone chaperone complexes, along with the reduced amounts of H3 histones in different cellular pools, strongly suggest that ATRX is implicated in H3 deposition and displays a histone chaperone function in Arabidopsis. In absence of Arabidopsis ATRX, we found no changes in H3 transcript levels but reduced non-nucleosomal and total histone pools, suggesting that upon ATRX loss, unbound histones are shuttled for degradation. Changes in histone availability and failure to correctly target histones to the site of deposition might then lead to reduced nucleosome occupancy. Alternatively, replication defects in *atrx* mutants may cause uncoupling of DNA replication and histone deposition, thereby affecting parental histone recycling and consequently nucleosome occupancy (Clynes et al., 2013).

Our genetic analysis revealed severe phenotypes in crosses with mutants for the HIR complex that is involved in replication-independent deposition (Tagami et al., 2004) and that interacts with H3.3 in Arabidopsis (Nie et al., 2014). In our study, we obtained viable *ubn2 atrx* or *cabin1 atrx* double mutants, in agreement with a previous study showing that these subunits of the complex are dispensable in certain contexts (Duc et al., 2015). In contrast, we rarely obtained *hira atrx* double mutants, and the surviving plants have severe developmental phenotypes. Therefore, in the absence of a functional HIR complex, ATRX may be required for an alternative pathway of H3.3 deposition. This is in agreement with the observation that Drosophila *hira* and *xnp* single mutants are viable and show moderate defects in H3.3 deposition (Schneiderman et al., 2012). The double mutants, however, die during larval development and are unable to assemble H3.3 into chromatin.

Indeed, when we looked specifically at the distribution of the canonical histone H3.1 in the different histone pools, we found that despite global reduction of H3 histone levels, the amount of H3.1 is globally unchanged. This suggests that the reduction of H3 in the non-nucleosomal and the total histone pools observed is mainly due to a reduction in H3.3, leading to an increase in the eH3.1/H4 ratio in *atrx* mutants compared to WT. At the nucleosomal level, we found that H3.1 enrichment at most heterochromatic sequences is maintained or even increased in *atrx* mutants, suggesting an alteration of the cellular H3.1/H3.3 balance.

At the chromatin level, the impact of ATRX loss can be clearly differentiated from the loss of HIRA. In *hira* mutants, nucleosome occupancy is affected at many genomic locations including active genes as well as transcriptionally repressed centromeric and pericentromeric repeats or transposons (this study, (Duc et al., 2015)). In contrast, we found that the *atrx* mutants show reduced nucleosomal occupancy only at few sites such as certain genes and the 45S rRNAs. This difference is also reflected in our genetic analysis. While the combination of *hir* and *caf* mutants results in lethality (Duc et al., 2015) showing that the HIR complex generally compensates impaired replication-coupled histone deposition, the *atrx fas2-5* double mutants are viable. Together, this implies that ATRX deposits histones in a more site-specific manner.

Within the ATRX-DAXX complex, DAXX binds histone H3.3 and determines the variant specificity of the hetero-complex (Elsässer et al., 2012; Liu et al., 2012) while ATRX guides the ATRX-DAXX hetero-complex to specific genomic sites through interaction with methylated histone tails. However, the mammalian ATRX does not bind directly to H3-H4 dimers. Indeed, no interaction between the Arabidopsis ATRX and histone H3 could be seen by Y2H assays (data not shown). To date, no DAXX homologue could be identified in the Arabidopsis genome based on protein sequence homology (this study, (Otero et al., 2014)), consistently with the absence of the mammalian DAXX-binding domain in the Arabidopsis ATRX. ATRX therefore might affect H3.3 deposition by interacting with a functional homologue of DAXX or yet unknown partners. Indeed, recent genome-wide analysis of the distribution of ATRX and DAXX proteins by ChIP-seq in embryonic stem cells revealed only a relatively small percentage of common binding sites (He et al., 2015), suggesting that ATRX and DAXX may have independent functions at certain targets. Efforts to identify ATRX binding partners will be an interesting avenue for a better understanding of ATRX function in Arabidopsis.

An alternative hypothesis that cannot be completely excluded is that ATRX affects H3 occupancy in an indirect manner, *e.g.* by affecting chromatin organization due to its remodeling activity or *via* changes in the enrichment in other core histones. Indeed, in mammals, ATRX co-immunoprecipitates with macroH2A, and ATRX loss leads to increased macroH2A enrichment at specific genomic sites (Ratnakumar et al., 2012).

In mammals, ATRX binds to repetitive satellite sequences and is involved in heterochromatin formation and silencing of certain transposable elements (He et al., 2015; Sadic et al., 2015; Voon et al., 2015). However, nucleosomal occupancy at centromeric or pericentromeric repetitive elements or enrichment in the plant heterochromatin signature H3K9me2 is not affected in *atrx* mutants. This is in agreement

with maintenance of transcriptional silencing at repetitive elements and at most of the transposons in the *atrx* mutant plants. The discrepancy in the observations between Arabidopsis and mice might be resolved by the fact that H3.3 is a mark for active chromatin and telomeres in Arabidopsis while centromeric and pericentromeric repeats are enriched in H3.1 (Shu et al., 2014; Vaquero-Sedas and Vega-Palas, 2013; Stroud et al., 2012; Wollmann et al., 2012). In contrast, in mouse embryonic stem cells, H3.3 is also found at pericentromeric domains (Drané et al., 2010; Goldberg et al., 2010; Wong et al., 2010) and enriched at specific classes of transposable elements or imprinted genes (Elsässer et al., 2015; He et al., 2015; Voon et al., 2015). Among the elements tested, only Mule (At2g15810) silencing was released in *atrx-1.* While it might result from an indirect effect at this specific genomic site, it may be of interest to analyze additional examples of DNA- and retro-transposons in *atrx* mutants.

Furthermore, the localization of ATRX to heterochromatin is determined by its ability to bind methylated H3K9 via its ADD domain (Iwase et al., 2012; Dhayalan et al., 2011; Eustermann et al., 2011). While we found that the overall structure of the ADD domain is conserved in Arabidopsis ATRX, the amino acids Y203, Y204 and Q219 critical for binding to H3K9me3 (Iwase et al., 2012) are not conserved (Supplemental Figure S1A). Therefore, it can also be envisaged that in plants ATRX may not be targeted specifically to heterochromatin, but may rather localize to regions with ongoing nucleosome displacement (Schneiderman et al., 2009) or bind similarly to the HIR complex directly to nucleosome-free gaps in DNA to direct histone deposition (Schneiderman et al., 2012; Ray-Gallet et al., 2011). Indeed, genome-wide studies identified many ATRX-binding sites outside of heterochromatin and at ribosomal genes (Law et al., 2010; He et al., 2015). Likewise, as shown here, changes in nucleosome occupancy in *atrx* rather situate at genes on chromosome arms (*HXK1* and the subtelomeric gene, At5g67640) as well as at ribosomal 45S rDNA loci, also located in subtelomeric regions.

Chromatin modifications are known to modulate ribosomal RNA expression level in various species such as yeast or mammals (Sandmeier et al., 2002; McStay and Grummt, 2008; Grummt and Längst, 2013). Indeed, in most eukaryotes, 45S rDNA genes are present in excess, resulting in the selective activation of only a subset of genes. This is in part controlled by epigenetic mechanisms involving histone modifications (Pontvianne et al., 2013, 2012; Preuss and Pikaard, 2007). Here, we showed that loss of ATRX leads to a reduced nucleosomal occupancy at 45S rDNA loci, while the 5S rDNA loci residing in the pericentromeric region of the chromosomes were unaffected. Globally, no change in eH3.1 enrichment at 45S rDNA was observed, implying that the reduction in nucleosome occupancy is mainly due to a reduced H3.3 incorporation. This should be confirmed in future when appropriate transgenic lines are available. Interestingly, H3K9me2 levels are globally increased at 45S rDNA, potentially as a means to compensate reduced nucleosome occupancy or differences in rDNA variant copy numbers.

The existence of 45S rDNA sequence variants in the Arabidopsis genome that differ by small indels (Pontvianne et al., 2010) allows to evidence changes in rDNA gene choice. Indeed, it has been previously reported that VAR1 is only weakly expressed

in WT plants even though it is most abundant at the DNA level. This is in agreement with higher nucleosome occupancy and H3K9 methylation when normalized to input, while VAR3 is rather depleted, which corroborates its expression profile. While the relative abundance of VAR1 rRNA as well as its enrichment in H3/H3K9me2 is unchanged in *atrx* mutants, expression shifts to increased VAR2 and reduced VAR3 rRNA in the mutant. This might be partly explained by increased VAR2 abundance in the *atrx-1* genome, reduced nucleosome occupancy and a lower H3K9me2 enrichment at this variant in the *atrx* mutant. Changes in variant abundance have previously been observed in different mutants, such as *nucleolin2* (*nuc2*) (Durut et al., 2014) or *atxr5 atxr6* (Pontvianne et al., 2012) and have been suggested to occur through gene conversion or selective re-replication coupled to intra-NOR recombination respectively. Other mutants such as those in the Arabidopsis CAF-1 complex progressively loose 45S rRNA genes (Muchová et al., 2015; Mozgová et al., 2010), however, without affecting the relative variant abundance (Pontvianne et al., 2013). Since 45S rRNA genes have been suggested to form complex DNA structures prone to replication defects (Muchová et al., 2015), we could speculate that replication defects in *atrx* are at the origin of the changes in relative 45S variant abundance.

The observed changes at the nucleosomal level are also reflected by changes in organization of the rDNA genes within the nucleus. More NORs are associated with the nucleolus, where the rRNA genes are actively transcribed, suggesting that previously silent VAR2 copies, residing on the normally silent NOR, are activated in *atrx-1* mutants.

Taken together, ATRX plays a major role in maintaining nucleosome occupancy at 45S loci, its loss affecting nucleosome occupancy and H3.1 enrichment in a variant-specific manner thereby modulating dosage and 45S rRNA gene choice. Hence, the Arabidopsis ATRX protein has similar properties to its mammalian and invertebrate counterparts and affects 45S rRNA expression in plants. This particular role of ATRX in 45S rDNA regulation is a yet unexplored aspect that might contribute to the complex disease phenotype of ATRX patients.

Material and Methods

Plant Material. Mutant *Arabidopsis* lines were obtained from the Nottingham Arabidopsis Stock Center (NASC) and/or were gifts from other laboratories. We used the following mutant *Arabidopsis* lines *atrx-1* (SALK_025687), *atrx-2* (SAIL-861-B04), *fas2-5* (SALK_147693) (Duc et al., 2015), *hira-1* (WiscDsLox362H05) (Ingouff et al., 2010), *ubn2-2* (GABI_018D02), cabin1-2 (SALK_099927), (Duc et al., 2015), the double mutant *atxr5* (SALK_130607) *atxr6* (SAIL_181_D09) (Jacob et al., 2009), *atr-2* (SALK_032841) (Rounds and Larsen, 2008), the triple mutant *ku80* (FLAG_DMT5) *xrcc1* (SALK_125373) *xpf* (N3819) (Charbonnel et al., 2011) kindly provided by S. Amiard (GReD). Except *xrcc1 ku80 xpf,* all mutants are in the Columbia background. Plants were grown on soil in a growth chamber under 16-h light/8-h dark cycles at 22ºC. The eH3.1 and eH3.3 lines were generated by transcriptional fusion of the genomic fragment containing the promoter and the genomic coding region of either *HTR9* or *HTR5* (stop codon excluded) with the FLAG-HA tag and the OCS (octopine synthase) terminator using classic cloning with restriction enzymes and the Gateway technology. Monolocus homozygous lines

were selected based on segregation of hygromycine resistance also encoded in the transgene, and further crosses to *atrx-1* were performed. For *in vitro* culture, seeds were surface-sterilized and sown on germination medium containing 0.8% w/v agar, 1% w/v sucrose and Murashige & Skoog salts (MS, M0255; Duchefa Biochemie, Netherlands). After 2d of stratification at 4ºC at dark, plants were grown under 16-h light/8-h dark cycles at 23ºC.

Phylogenetic analysis. To identify the ATRX Arabidopsis ortholog, we performed interspecies Blast searches with the mammalian protein sequence. Conserved domains were aligned using the program Muscle with the ESPript (Easy Sequencing in PostScript) program (Robert and Gouet, 2014) with default settings. Trees were constructed from amino acid sequences of the conserved motifs. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992). The tree with the highest log likelihood (-15825.7621) was used. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Distance bootstrap analyses consisted of 1,000 replicates. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with a superior log likelihood value. The tree is drawn to scale, with branch lengths measured in number of substitutions per site. The analysis involved 11 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 1002 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013).

Protein structure homology model. To establish a model for the ADD domain, we generated a three-dimensional structural model of human and Arabidopsis ATRX. The CPH models-3.2 server (www.cbs.dtu.dk/services/CPHmodels) was used to generate the ribbon protein models and the Chimera software (www.cgl.ucsf.edu/chimera) for the protein model superimposition.

Documentation of phenotypes. Images of dissected siliques and flowers were taken using a Leica binocular and the LAS 3.6 Software (Leica Microsystems, Switzerland), with a 0.63x and a 2x magnification, respectively. Viability of mature pollen grains was assayed as described in (Alexander, 1969). Anthers from stained flowers were isolated and photographed using a Zeiss Axioplan microscope and the Axiovision 4.2 software (Carl Zeiss Vision GmbH, Le Pecq, France). For seed counting, we distinguished viable green seeds from unfertilized ovules as described in (Duc et al., 2015). For rosette area measurement, digital photographs of double mutants and their respective single mutant sister plants were processed with the ImageJ (Fiji) software to measure rosette area (total area of rosette leaves).

RNA extraction and RT-PCR. RNAs were extracted with Tri-Reagent (Euromedex) according to manufacturer's instructions, then treated with RQ1 DNase I (Promega) and purified with phenol-chloroform extraction. Reverse transcription was done either with oligo(dT)15 or with random hexamers using M-MLV reverse transcriptase (Promega). For analysis of *ATRX* expression in *atrx* alleles, reverse-transcription was done with the ATRX_RT_Rev primer (5'-GGGACCCGTTGAACTTCCTCCC-3')

combined with random hexamers. Obtained cDNAs were diluted 1:3 and used in PCR (Promega Flexi) or in quantitative PCR with the LightCycler® 480 SYBR Green I Master kit on the Roche LightCycler® 480. Transcript levels of interest were normalized to *At2g28390* (Czechowski et al., 2005) or to *UBC28* (At1g64230) for Supplemental Figure S2 using the comparative threshold cycle method. RT-qPCR histograms show means of transcript levels ± SEM obtained for two independent PCR amplifications of three biological replicates. The y-axis shows the fold change relative to WT (WT set to 1) after normalization to *At2g28390* expression. For analysis of silencing release, the following loci were tested: Ta3 and Athila LTR pericentromeric retrotransposons, Mule transposons At1g40097 and At1g43280, the Mu1 (At4g08680) transposons and SINE retrotransposon AtSN1 (At3TE63860) (Vaillant et al., 2006), the CACTA1 transposon and the T5L23.26 CACTA-like transposon (located at the heterochromatin knob of chromosome 4) (Ono et al., 2006) and the COPIA78 LTR transposon (Pecinka et al., 2010).

Analysis of 45S rDNA and rRNA variants. For analysis of 45S rDNA expression, cDNA synthesis was performed on 2.5 µg of DNaseI-treated RNA with the 3allrRNAvar primer using M-MLV reverse transcriptase (Promega). Obtained cDNAs were diluted 1:2 and used in PCR (Promega Flexi) with 25 to 40 cycles. For the relative abundance of each class of rDNA and rRNA variants, PCR was performed with 5allrRNAvar/3allrRNAvar primers. Quantification of variants was performed on non-saturated signals using Multi Gauge software (Fujfilm).

ChIP analysis. 2.5-week-old plantlets were formaldehyde cross-linked and chromatin immunoprecipitation carried out as previously described (Bowler et al., 2004) with minor modifications: chromatin was sheared with the Diagenode Bioruptor (10 cycles of 30s ON and 1.5 min OFF). Protein A-coupled Dynabeads (Invitrogen) or Anti-FLAG® M2 Magnetic beads (Sigma) were used, and the sonicated chromatin was pre-cleared in presence of Invitrogen magnetic beads for 3h, before immuno-precipitation with the anti-H3 antibody (Abcam, ab1791), with the anti-H3K9me2 antibody (Abcam, ab1220) or with the Anti-FLAG® M2 Magnetic beads (Sigma). DNA was quantified using qPCR and normalized relatively to input.

Protein Extraction and Western Blot Analysis. To recover the non-nucleosomal fraction of histones present in the cytoplasm and in the nucleoplasm, proteins were extracted from 100 mg of plantlets according to (Durut et al., 2014). For the total extract corresponding to the histones present in the cytoplasm, nucleoplasm and chromatin and the nucleus, proteins were recovered from 1g of plantlets with Honda buffer (Ficoll 400 2.5%; Dextran T40 5%; Sucrose 0.4M; Tris-HCl 25mM pH7.4; MgCl2 10mM; beta-mercaptoethanol 10mM; PMSF 0.5mM; complete Protease Inhibitor Cocktail Tablets Roche) (Honda et al., 1966). Briefly, grinded tissues were homogenized in 15mL of Honda buffer and filtered through a double layer of Miracloth (Millipore). After removal of an aliquot that constitutes the “total” extract, Triton X-100 was added at a 0.5% concentration and samples were submitted to rotation on a wheel at cold for 15min. After 5min centrifugation at 1,500g, pellets were washed with Honda buffer containing 0.1% Triton X-100. These nuclei-rich preparations were centrifuged at 1,500g for 5min, and recovered pellets were washed with Honda buffer without Triton X-100. Nuclei were

recovered after 5min centrifugation at 1,500g to constitute the “nuclear” fraction. Western blots were probed with the anti-H3 antibody (Abcam, ab1791, 1/3,000), anti-HA antibody (Abcam, ab9110, 1/1,000) or with the anti-H4 antibody (Abcam, ab10158, 1/1,000). Equal loading was confirmed with an anti-actin antibody (Sigma; 1/3,000). Primary antibodies were revealed by incubation with anti-rabbit (1/3,000) or anti-mouse (1/3,000; Abliance) secondary antibodies. Immunoblot chemiluminescence was revealed using ECL western blotting detection reagents (GE Healthcare Bio-Sciences). Densitometric analysis of immunoreactive protein bands was performed on non-saturated signals using Multi Gauge software (Fujfilm) and H3, H4 and HA levels normalized to actin with WT set to 100%.

Fluorescence *in situ* hybridization. For nuclear spreads, rosette leaves from 2-week-old *in vitro*–grown plants were fixed in ethanol-acetic acid (3:1 v/v). FISH was performed essentially as described (Probst et al., 2003). Biotin-labeled 45S rDNA probes were generated by nick-translation (Roche). Slides were analyzed with the Leica MMAF microscope and images processed with ImageJ.

Primers. All primer sequences are listed in Supplemental Table S4.

Acknowledgments

We thank S. Amiard for seeds and help with the EdU detection, HU and gamma-irradiation experiments, O. Mittelsten Scheid for critical reading of the manuscript and F. Pontvianne for valuable advice with 45S rDNA variant analysis and for critical reading of the manuscript.

Author contributions

CD and AVP designed the research; CD, GD, LS, MB, SLG, SC and AVP performed research; CD, AP, CT and AVP analyzed the data; AP and AVP contributed reagents/materials/analysis tools; CD, CT an AVP wrote the paper.

References:

Adachi, S. et al. (2011). Programmed induction of endoreduplication by DNA double-strand breaks in Arabidopsis. Proc. Natl. Acad. Sci. 108 : 10004–10009.

Alekseyenko, A.A., Gorchakov, A.A., Zee, B.M., Fuchs, S.M., Kharchenko, P. V, and Kuroda, M.I. (2014). Heterochromatin-associated interactions of Drosophila HP1a with dADD1 , HIPP1 , and repetitive RNAs. Genes Dev. 28: 1445–1460.

Alexander, M. (1969). Differential staining of aborted and nonaborted pollen. Stain Technol. 44: 117–22.

Alexeev, A., Mazin, A., and Kowalczykowski, S.C. (2003). Rad54 protein possesses chromatin-remodeling activity stimulated by the Rad51-ssDNA nucleoprotein filament. Nat. Struct. Biol. 10: 182–186.

Amiard, S., Charbonnel, C., Allain, E., Depeiges, A., White, C.I., and Gallego, M.E. (2010). Distinct Roles of the ATR Kinase and the Mre11-Rad50-Nbs1 Complex in the Maintenance of Chromosomal Stability in Arabidopsis. Plant Cell 22: 3020–3033.

Bassett, A.R., Cooper, S.E., Ragab, A., and Travers, A. a. (2008). The chromatin remodelling factor dATRX is involved in heterochromatin formation. PLoS One 3: e2099.

Benoit, M., Layat, E., Tourmente, S., and Probst, A. V (2013). Heterochromatin dynamics during developmental transitions in Arabidopsis - a focus on ribosomal DNA loci. Gene 526: 39–45.

Bernatavichute, Y. V, Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-Wide Association of Histone H3 Lysine Nine Methylation with CHG DNA Methylation in Arabidopsis thaliana. PLoS One 3: e3156.

Bowler, C., Benvenuto, G., Laflamme, P., Molino, D., Probst, A. V, Tariq, M., and Paszkowski, J. (2004). Chromatin techniques for plant cells. Plant J. 39: 776–789.

Campell, B.R., Song, Y., Posch, T.E., Cullis, C.A., and Town, C.D. (1992). Sequence and organization of 5S ribosomal RNA-encoding genes of Arabidopsis thaliana. Gene 112: 225–228.

Charbonnel, C., Allain, E., Gallego, M.E., and White, C.I. (2011). Kinetic analysis of DNA double-strand break repair pathways in Arabidopsis. DNA Repair (Amst). 10: 611–9.

Cloix, C., Tutois, S., Mathieu, O., Cuvillier, C., Espagnol, M.C., Picard, G., and Tourmente, S. (2000). Analysis of 5S rDNA Arrays in Arabidopsis thaliana : Physical Mapping and Chromosome-Specific Polymorphisms Analysis of 5S rDNA Arrays in Arabidopsis thaliana : Physical Mapping and Chromosome-Specific Polymorphisms. 10: 679–690.

Clynes, D., Higgs, D.R., and Gibbons, R.J. (2013). The chromatin remodeller ATRX: A

repeat offender in human disease. Trends Biochem. Sci. 38: 461–466.

Collins, N., Poot, R.A., Kukimoto, I., Garcia-Jimenez, C., Dellaire, G., and Varga-Weisz, P.D. (2002). An ACF1-ISWI chromatin-remodeling complex is required for DNA replication through heterochromatin. Nat Genet 32: 627–632.

Culligan, K., Tissier, A., and Britt, A. (2004). ATR Regulates a G2-Phase Cell-Cycle Checkpoint in Arabidopsis thaliana. Plant Cell 16: 1091–1104.

Czechowski, T., Stitt, M., Altmann, T., Udvardi, M.K., and Scheible, W.-R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiol. 139: 5–17.

Dhayalan, A., Tamas, R., Bock, I., Tattermusch, A., Dimitrova, E., Kudithipudi, S., Ragozin, S., and Jeltsch, A. (2011). The ATRX-ADD domain binds to H3 tail peptides and reads the combined methylation state of K4 and K9. Hum. Mol. Genet. 20: 2195–2203.

Drané, P., Ouararhni, K., Depaux, A., Shuaib, M., and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev. 24: 1253–1265.

Duc, C., Benoit, M., Goff, S. Le, Simon, L., Poulet, A., Cotterell, S., Tatout, C., and Probst, A. V. (2015). The histone chaperone complex HIR maintains nucleosome occupancy and counterbalances impaired histone deposition in CAF-1 mutants. Plant J. 81: 707–22.

Durut, N. et al. (2014). A Duplicated NUCLEOLIN Gene with Antagonistic Activity Is Required for Chromatin Organization of Silent 45S rDNA in Arabidopsis. Plant Cell 1: 1–16.

Elsässer, S.J., Huang, H., Lewis, P.W., Chin, J.W., Allis, C.D., and Patel, D.J. (2012). DAXX envelops a histone H3.3–H4 dimer for H3.3-specific recognition. Nature 491: 560–565.

Elsässer, S.J., Noh, K.-M., Diaz, N., Allis, C.D., and Banaszynski, L. a. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. Nature 522: 240–4.

Emelyanov, A. V., Konev, A.Y., Vershilova, E., and Fyodorov, D. V. (2010). Protein complex of Drosophila ATRX/XNP and HP1a is required for the formation of pericentric beta-heterochromatin in vivo. J. Biol. Chem. 285: 15027–15037.

Eustermann, S., Yang, J.-C., Law, M.J., Amos, R., Chapman, L.M., Jelinska, C., Garrick, D., Clynes, D., Gibbons, R.J., Rhodes, D., Higgs, D.R., and Neuhaus, D. (2011). Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. Nat. Struct. Mol. Biol. 18: 777–782.

Filipescu, D., Szenker, E., and Almouzni, G. (2013). Developmental roles of histone H3 variants and their chaperones. Trends Genet. 29: 630–640.

Fransz, P., Armstrong, S., Alonso-Blanco, C., Fischer, T.C., Torres-Ruiz, R. a., and Jones, G. (1998). Cytogenetics for the model system Arabidopsis thaliana. Plant J. 13: 867–876.

Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P., and Firoozabady, E. (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science 220: 1049–51.

Garrick, D., Sharpe, J.A., Arkell, R., Dobbie, L., Smith, A.J.H., Wood, W.G., Higgs, D.R., and Gibbons, R.J. (2006). Loss of Atrx Affects Trophoblast Development and the Pattern of X-Inactivation in Extraembryonic Tissues. PLoS Genet 2: e58.

Gibbons, R.J., Picketts, D.J., and Higgs, D.R. (1995a). Syndromal mental retardation due to mutations in a regulator of gene expression. Hum. Mol. Genet. 4: 1705–1709.

Gibbons, R.J., Picketts, D.J., Villard, L., and Higgs, D.R. (1995b). Mutations in a putative global transcriptional regulator cause X-linked mental retardation with α-thalassemia (ATR-X syndrome). Cell 80: 837–845.

Gokhman, D., Livyatan, I., Sailaja, B.S., Melcer, S., and Meshorer, E. (2013). Multilayered chromatin analysis reveals E2f, Smad and Zfx as transcriptional regulators of histones. Nat. Struct. Mol. Biol. 20: 119–26.

Goldberg, A.D. et al. (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell 140: 678–691.

Grummt, I. and Längst, G. (2013). Epigenetic control of RNA polymerase I transcription in mammalian cells. Biochim. Biophys. Acta - Gene Regul. Mech. 1829: 393–404.

Hayashi, K., Hasegawa, J., and Matsunaga, S. (2013). The boundary of the meristematic and elongation zones in roots: endoreduplication precedes rapid cell expansion. Sci. Rep. 3: 2723.

He, Q., Kim, H., Huang, R., Lu, W., Tang, M., Shi, F., Yang, D., Zhang, X., Huang, J., Liu, D., and Songyang, Z. (2015). The Daxx/Atrx Complex Protects Tandem Repetitive Elements during DNA Hypomethylation by Promoting H3K9 Trimethylation. Cell Stem Cell 17: 273–286.

Honda, S.I., Hongladarom, T., and Laties, G.G. (1966). A new isolation medium for plant organelles. J. Exp. Bot. 17: 460–472.

Huh, M.S., Price O’Dea, T., Ouazia, D., McKay, B.C., Parise, G., Parks, R.J., Rudnicki, M.A., and Picketts, D.J. (2012). Compromised genomic integrity impedes muscle growth after Atrx inactivation . J. Clin. Invest. 122: 4412–4423.

Ingouff, M. et al. (2010). Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in arabidopsis. Curr. Biol. 20: 2137–2143.

Iwase, S., Xiang, B., Ghosh, S., Ren, T., Lewis, P.W., Cochrane, J.C., Allis, C.D., Picketts, D.J., Patel, D.J., Li, H., and Shi, Y. (2012). ATRX links atypical histone methylation recognition mechanisms to human brain function. Nat. Struct. Mol. Biol. 18: 769–776.

Jacob, Y., Feng, S., LeBlanc, C.A., Bernatavichute, Y. V, Stroud, H., Cokus, S., Johnson, L.M., Pellegrini, M., Jacobsen, S.E., and Michaels, S.D. (2009). ATXR5 and ATXR6 are novel H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat. Struct. Mol. Biol. 16: 763–768.

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992). The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. CABIOS 8 : 275–282.

Kaya, H., Shibahara, K.I., Taoka, K.I., Iwabuchi, M., Stillman, B., and Araki, T. (2001). FASCIATA genes for chromatin assembly factor-1 in Arabidopsis maintain the cellular organization of apical meristems. Cell 104: 131–142.

Konev, A.Y., Tribus, M., Park, S.Y., Podhraski, V., Lim, C.Y., Emelyanov, A. V, Vershilova, E., Pirrotta, V., Kadonaga, J.T., Lusser, A., and Fyodorov, D. V (2007). CHD1 Motor Protein Is Required for Deposition of Histone Variant H3.3 into Chromatin in Vivo. Sci. 317 : 1087–1090.

De La Fuente, R., Baumann, C., and Viveiros, M.M. (2015). ATRX contributes to epigenetic asymmetry and silencing of major satellite transcripts in the maternal genome of the mouse embryo. Development 142: 1–12.

Law, M.J. et al. (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. Cell 143: 367–78.

Lee, N.G., Hong, Y.K., Yu, S.Y., Han, S.Y., Geum, D., and Cho, K.S. (2007). dXNP, a Drosophila homolog of XNP/ATRX, induces apoptosis via Jun-N-terminal kinase activation. FEBS Lett. 581: 2625–2632.

Leung, J.W.-C., Ghosal, G., Wang, W., Shen, X., Wang, J., Li, L., and Chen, J. (2013). Alpha Thalassemia/Mental Retardation Syndrome X-linked Gene Product ATRX Is Required for Proper Replication Restart and Cellular Resistance to Replication Stress. J. Biol. Chem. 288: 6342–6350.

Lewis, P.W., Elsaesser, S.J., Noh, K.-M., Stadler, S.C., and Allis, C.D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proc. Natl. Acad. Sci. U. S. A. 107: 14075–14080.

Liu, C.-P., Xiong, C., Wang, M., Yu, Z., Yang, N., Chen, P., Zhang, Z., Li, G., and Xu, R.-M. (2012). Structure of the variant histone H3.3-H4 heterodimer in complex with its chaperone DAXX. Nat. Struct. Mol. Biol. 19: 1287–1292.

López-Falcón, B., Meyer-Nava, S., Hernández-Rodríguez, B., Campos, A., Montero, D., Rudiño, E., Vázquez, M., Zurita, M., and Valadez-Graham, V. (2014). Characterization

of the Drosophila Group Ortholog to the Amino-Terminus of the Alpha-Thalassemia and Mental Retardation X-Linked (ATRX) Vertebrate Protein. PLoS One 9: e113182.

McDowell, T.L. et al. (1999). Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. Proc. Natl. Acad. Sci. U. S. A. 96: 13983–13988.

McStay, B. and Grummt, I. (2008). The Epigenetics of rRNA Genes: From Molecular to Chromosome Biology. Annu. Rev. Cell Dev. Biol. 24: 131–157.

Mizuguchi, G., Shen, X., Landry, J., Wu, W.-H., Sen, S., and Wu, C. (2004). ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. Sci. 303 : 343–348.

Mozgová, I., Mokros, P., and Fajkus, J. (2010). Dysfunction of chromatin assembly factor 1 induces shortening of telomeres and loss of 45S rDNA in Arabidopsis thaliana. Plant Cell 22: 2768–2780.

Muchová, V., Amiard, S., Mozgová, I., Dvořáčková, M., Gallego, M.E., White, C., and Fajkus, J. (2015). Homology-dependent repair is involved in 45S rDNA loss in plant CAF-1 mutants. Plant J. 81: 198–209.

Nie, X., Wang, H., Li, J., Holec, S., and Berger, F. (2014). The HIRA complex that deposits the histone H3.3 is conserved in Arabidopsis and facilitates transcriptional dynamics. Biol. Open: 1–9.

Okada, T., Endo, M., Singh, M.B., and Bhalla, P.L. (2005). Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3. Plant J. 44: 557–568.

Ono, T., Kaya, H., Takeda, S., Abe, M., Ogawa, Y., Kato, M., Kakutani, T., Scheid, O.M., Araki, T., and Shibahara, K.I. (2006). Chromatin assembly factor 1 ensures the stable maintenance of silent chromatin states in Arabidopsis. Genes to Cells 11: 153–162.

Otero, S., Desvoyes, B., and Gutierrez, C. (2014). Histone h3 dynamics in plant cell cycle and development. Cytogenet. Genome Res. 143: 114–24.

Pchelintsev, N., McBryan, T., Rai, T., van Tuyn, J., Ray-Gallet, D., Almouzni, G., and Adams, P. (2013). Placing the HIRA histone chaperone complex in the chromatin landscape. Cell Rep. 3: 1012–1019.

Pecinka, A., Dinh, H.Q., Baubec, T., Rosa, M., Lettner, N., and Mittelsten Scheid, O. (2010). Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. Plant Cell 22: 3118–3129.

Pontvianne, F. et al. (2010). Nucleolin Is Required for DNA Methylation State and the Expression of rRNA Gene Variants in Arabidopsis thaliana. PLoS Genet. 6: e1001225.

Pontvianne, F., Blevins, T., Chandrasekhara, C., Feng, W., Stroud, H., Jacobsen, S.E., Michaels, S.D., and Pikaard, C.S. (2012). Histone methyltransferases regulating rRNA gene dose and dosage control in Arabidopsis. Genes Dev. 26: 945–957.

Pontvianne, F., Blevins, T., Chandrasekhara, C., Mozgova, I., Hassel, C., Pontes, O.M.F., Tucker, S., Mokros, P., Muchova, V., Fajkus, J., and Pikaard, C.S. (2013). Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. Genes Dev. 27: 1545–1550.

Pontvianne, F., Matia, I., Douet, J., Tourmente, S., Medina, F.J., Echeverria, M., and Saez-Vasquez, J. (2007). Characterization of AtNUC-L1 reveals a central role of nucleolin in nucleolus organization and silencing of AtNUC-L2 gene in Arabidopsis. Mol. Biol. Cell 18: 369–379.

Preuss, S. and Pikaard, C.S. (2007). rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. Biochim. Biophys. Acta 1769: 383–392.

Probst, A. V, Fransz, P.F., Paszkowski, J., and Scheid, O.M. (2003). Two means of transcriptional reactivation within heterochromatin. Plant J. 33: 743–749.

Ramirez-Parra, E. and Gutierrez, C. (2007). E2F Regulates FASCIATA1, a Chromatin Assembly Gene Whose Loss Switches on the Endocycle and Activates Gene Expression by Changing the Epigenetic Status. Plant Physiol. 144: 105–120.

Ratnakumar, K. et al. (2012). ATRX-mediated chromatin association of histone variant macroH2A1 regulates α-globin expression. Genes Dev. 26: 433–438.

Ray-Gallet, D., Woolfe, A., Vassias, I., Pellentz, C., Lacoste, N., Puri, A., Schultz, D.C., Pchelintsev, N. a, Adams, P.D., Jansen, L.E.T., and Almouzni, G. (2011). Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. Mol. Cell 44: 928–41.

Robert, X. and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42: W320–W324.

Rounds, M.A. and Larsen, P.B. (2008). Aluminum-dependent root-growth inhibition in Arabidopsis results from AtATR-regulated cell-cycle arrest. Curr. Biol. 18: 1495–500.

Sadic, D., Schmidt, K., Groh, S., Kondofersky, I., Ellwart, J., Fuchs, C., Theis, F.J., and Schotta, G. (2015). Atrx promotes heterochromatin formation at retrotransposons. EMBO Rep. 16: 836–850.

Sandmeier, J.J., French, S., Osheim, Y., Cheung, W.L., Gallo, C.M., Beyer, A.L., and Smith, J.S. (2002). RPD3 is required for the inactivation of yeast ribosomal DNA genes in stationary phase. EMBO J. 21: 4959–4968.

Sawatsubashi, S. et al. (2010). A histone chaperone, DEK, transcriptionally coactivates a nuclear receptor. Genes Dev. 24: 159–170.

Schneiderman, J.I., Orsi, G.A., Hughes, K.T., Loppin, B., and Ahmad, K. (2012). Nucleosome-depleted chromatin gaps recruit assembly factors for the H3.3 histone variant. Proc. Natl. Acad. Sci. U. S. A. 109: 19721–19726.

Schneiderman, J.I., Sakai, A., Goldstein, S., and Ahmad, K. (2009). The XNP remodeler targets dynamic chromatin in Drosophila. Proc. Natl. Acad. Sci. U. S. A. 106: 14472–14477.

Schönrock, N., Exner, V., Probst, A., Gruissem, W., and Hennig, L. (2006). Functional genomic analysis of CAF-1 mutants in Arabidopsis thaliana. J. Biol. Chem. 281: 9560–9568.

Shaked, H., Avivi-Ragolsky, N., and Levy, A. a. (2006). Involvement of the arabidopsis SWI2/SNF2 chromatin remodeling gene family in DNA damage response and recombination. Genetics 173: 985–994.

Shu, H., Nakamura, M., Siretskiy, A., Borghi, L., Moraes, I., Wildhaber, T., Gruissem, W., and Hennig, L. (2014). Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes. Genome Biol. 15: R62.

Smith, S. and Stillman, B. (1989). Purification and characterization of CAF-I, a human cell factor required for chromatin assembly during DNA replication in vitro. Cell 58: 15–25.

Steimer, a, Amedeo, P., Afsar, K., Fransz, P., Mittelsten Scheid, O., and Paszkowski, J. (2000). Endogenous targets of transcriptional gene silencing in Arabidopsis. Plant Cell 12: 1165–1178.

Stroud, H., Otero, S., Desvoyes, B., Ramírez-parra, E., and Jacobsen, S.E. (2012). Genome-wide analysis of histone H3.1 and H3.3 variants in Arabidopsis thaliana. Proc. Natl. Acad. Sci. 109: 5370–5375.

Tagami, H., Ray-Gallet, D., Almouzni, G., and Nakatani, Y. (2004). Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. Cell 116: 51–61.

Talbert, P. et al. (2012). A unified phylogeny-based nomenclature for histone variants. Epigenetics Chromatin 5: 7.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30: 2725–2729.

Tang, J., Wu, S., Liu, H., Stratt, R., Barak, O.G., Shiekhattar, R., Picketts, D.J., and Yang, X.

(2004). A Novel Transcription Regulatory Complex Containing Death Domain-associated Protein and the ATR-X Syndrome Protein. J. Biol. Chem. 279: 20369–20377.

Thompson, H.L., Schmidt, R., and Dean, C. (1996). Identification and distribution of seven classes of middle-repetitive DNA in the Arabidopsis thaliana genome. Nucleic Acids Res. 24: 3017–3022.

Ülker, B., Peiter, E., Dixon, D.P., Moffat, C., Capper, R., Bouché, N., Edwards, R., Sanders, D., Knight, H., and Knight, M.R. (2008). Getting the most out of publicly available T-DNA insertion lines. Plant J. 56: 665–677.

Vaillant, I., Schubert, I., Tourmente, S., and Mathieu, O. (2006). MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis. EMBO Rep. 7: 1273–8.

Vaquero-Sedas, M.I. and Vega-Palas, M. a (2013). Differential association of Arabidopsis telomeres and centromeres with histone H3 variants. Sci. Rep. 3: 1202.

Voon, H.P.J., Hughes, J.R., Rode, C., De La Rosa-Velázquez, I.A., Jenuwein, T., Feil, R., Higgs, D.R., and Gibbons, R.J. (2015). ATRX Plays a Key Role in Maintaining Silencing at Interstitial Heterochromatic Loci and Imprinted Genes. Cell Rep. 11: 1–14.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1: 945–951.

Wollmann, H., Holec, S., Alden, K., Clarke, N.D., Jacques, P.É., and Berger, F. (2012). Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the Arabidopsis transcriptome. PLoS Genet. 8: 28–31.

Wong, L.H., Mcghie, J.D., Sim, M., Anderson, M. a, Ahn, S., Hannan, R.D., George, A.J., Morgan, K. a, Mann, J.R., and Choo, K.H.A. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. Genome Res. 20: 351–360.

Xue, Y., Gibbons, R., Yan, Z., Yang, D., McDowell, T.L., Sechi, S., Qin, J., Zhou, S., Higgs, D., and Wang, W. (2003). The ATRX syndrome protein forms a chromatin-remodeling complex with Daxx and localizes in promyelocytic leukemia nuclear bodies. Proc. Natl. Acad. Sci. U. S. A. 100: 10635–10640.

Supplemental Information

Supplemental Material and Methods

Gamma irradiation. Seeds were sown on the above-described MS medium. After stratification, plates were irradiated with gamma rays for 30min at a dose rate of 3Gy/min (100Gy in total) and then *in vitro* grown for 7d. Seedlings with true leaves were scored in a double blind assay. We calculated the ratio in percentage of plants with true leaves after irradiation relatively to the number of plants with true leaves without irradiation. This calculation separates the effect of the DNA-damaging treatment from the vigor of the *atrx* mutants.

Ploidy Analysis. Nuclei were prepared from 2-week-old *in vitro*-grown cotyledons with a modified version of the original Galbraith method (Galbraith et al., 1983). For ploidy analysis of each genotype, three independent preparations of pooled cotyledon material from 50 plants were used. Cytometric analysis was carried out using the *Attune® Acoustic* Focusing cytometer.

Hydroxyurea treatment. Root length of seedlings grown in absence or presence of 1mM hydroxyurea (HU, Sigma) was measured with ImageJ (Fiji) software after 2, 5 and 7d using the “straight line” tool. Since roots are shorter under control conditions in both *atrx-1* and *atrx-2* alleles compared to WT, we calculated the percentage of root length in presence of HU relative to the length under control growth conditions.

EdU treatement. Arabidopsis seedlings were germinated on MS plates and after 7d were transferred to liquid medium containing 10µM of EdU for 1 hour. Seedlings were then rinsed twice in liquid MS medium and fixed in 3.7% formaldehyde in 1X PBS. Root tips were digested for 1 h at 37ºC in a 1% (w/v) cellulase, 0.5% (w/v) cytohelicase, 1% (w/v) pectolyase (Sigma-Aldrich; Refs. C1794, C8274, P5936), squashed and quickly frozen in liquid nitrogen. Apexes were then permeabilized in 0.5% Triton X-100 and EdU detection was performed with the Invitrogen Click-iT EdU Alexa Fluor 488 Imaging kit (Amiard et al., 2010) according to manufacturer’s instructions. Nuclei with early versus late replicating patterns were distinguished according to (Hayashi et al., 2013) for about 700 nuclei derived from at least six root tips per genotype.

Supplemental Figures Legends

Supplemental Figure S1. Evolutionary conservation of the ATRX chaperones and expression of Arabidopsis *ATRX*.

(A) Detailed amino-acid sequence alignment of the ADD domain shows divergence in amino acids required for recognition of H3K9me3 in HsATRX (blue triangles), but reveals highly conserved cysteines (red stars). Identical residues were shaded in red and similar residues in yellow.

(B) Detailed amino-acid sequence alignment of the DExDc domain shows conservation of the consensus amino acid sequences (GXGKT) for the Walker A motif (Walker et al., 1982), highly conserved in many ATPases including helicases. The GXGKT consensus is displayed by green stars. Identical residues were shaded in red and similar residues in yellow.

(C) Detailed amino-acid sequence alignment of the HELICc domain. Identical residues were shaded in red and similar residues in yellow.

(D) Ribbon representation of AtATRX (left) and HsATRX (middle) ADD domains and their superimposition (right). The ADD domain contains both a GATA-like zinc finger (green) and a PHD finger (red). The CPH models-3.2 server was used to create the models and the Chimera software for superimposition, AtATRX being displayed in blue.

(E) *ATRX* expression in WT cotyledons, rosette, cauline leaves and inflorescences revealed by qRT-PCR.

Supplemental Figure S2. Analysis of *atrx* mutant alleles.

(A) Gene structure of the Arabidopsis *ATRX*. Exons, black rectangles; UTRs, purple rectangles; introns, lines; T-DNA insertion, triangle; LB, left border. The amplified regions used in (B) are displayed by green lines.

(B) Analysis of ATRX transcripts in the *atrx-1* and *atrx-2* mutants by semi-quantitative RT-PCR. The amplified regions are displayed in (A). The *atrx-2* mutant is a SAIL line for which the T-DNA vector contains a bidirectional 1′2′ promoter which can lead to activation of gene transcription from flanking sequences (Ülker et al., 2008). Hence, expression detected by qRT-PCR could either be driven by this 1′2′ promoter or result from transcription coupled with a splicing event deleting the region between exon 9 and exon 21, causing leaky expression of *ATRX*.

(C) Location of the *atrx* T-DNA insertions relative to *ATRX* transcripts and ATRX protein functional domains.

(D) Ploidy level distribution in nuclei of *atrx-1* mutant cotyledons from 2-week-old plantlets. Student test, *, $p<0.05$.

(E) Flow cytometry profiles of WT, *atrx-1* and *atrx-2* 2-week-old cotyledons. The number above each peak (CV, coefficient of variation) indicates the number of fluorescence intensity units that enclose the central 68% of nuclei for that endoreplication level.

(F) Root growth of *atrx-1, atrx-2* and *atr-2* mutant seedlings in presence of hydroxyurea (HU). The y-axis shows the root length ratio of plants for each treated population in relation to root length of plants in mock conditions separating the effect of the HU treatment from the innate vigor of the mutants. Root growth is completely arrested in *atr-2* mutants (Culligan et al., 2004) so that no bar is visible in the histogram.

(G) Frequency of nuclei with early and late replication patterns revealed by EdU incorporation. Early and late S phase nuclei were distinguished as in (Hayashi et al., 2013). Student test, ***, $p<0.001$.

(H) True leaf assay for gamma-irradiation (100Gy) treatment. The percentage of 7d-old plants with true leaves from irradiated seeds relative to non-irradiated seeds is presented. The triple mutant *ku80 xrcc1 xpf* (Charbonnel et al., 2011) affected in DNA damage repair was used as a control. In this mutant, the accumulation of DNA damage causes cell cycle arrest in the apical meristem and impairs the emergence of the first true leaves. Student test, *, $p<0.05$; **, $p<0.01$.

Supplemental Figure S3. Genetic interaction between CAF-1 and HIR histone chaperones and ATRX.

(A-B) Representative flowers of F3 sister plants obtained from crosses between *hira-1* and *atrx* mutants. Some *hira-1/hira atrx-1* displayed shorter filaments (A, red arrows).

(C-D) Pollen viability assessed by Alexander staining on F3 sister plants. The *hira-1 atrx-1/ATRX* anthers contain more non-viable pollen (green color) indicated by white arrows (C). The *hira-1 atrx-2/ATRX* anthers have reduced pollen content and contain more non-viable pollen (D).

(E) Quantification of seed content in F3 sister plant siliques obtained from crosses between *hira-1* and *atrx* mutants. Quantification is based on at least 30 siliques pooled from 3 plants.

(F-G) Representative 3-week-old F3 plants derived from crosses of *atrx-2* alleles with *ubn2-2* (F) or *cabin1-2* (G) grown on soil.

(H) Quantification of total rosette surface area of the F3 sister plants described in (F) and (G). Quantification is based on at least six 2-week-old plants for each genotype. Student test, **, $p<0.01$; ***, $p<0.001$.

(I-J) Representative flowers of F2 sister plants from crosses between *fas2-5* and *atrx* mutants. Plants are F2 sister plants derived from *atrx-1* (I) and *atrx-2* (J) crosses with *fas2-5*.

(K-L) Pollen viability assessed by Alexander staining of anthers from F2 sister plants obtained from crossed between *fas2-5* and *atrx-1* (K) or *atrx-2* (L). The *fas2-5 atrx-2* anthers have strongly reduced pollen content and increased amounts of non-viable pollen (L).

(M) Quantification of seed content in *FAS2 ATRX*, *fas2-5 ATRX* and *fas2-5 atrx-2/ATRX* mutant siliques. Quantification is based on at least 30 siliques pooled from 4 plants.

(N) Representative dissected siliques from *fas2-5 atrx-2* double mutants. Siliques contain either no seed (left silique) or one single seed (right silique).

Supplemental Figure S4. Effects of ATRX loss on H3K9me2 enrichment and expression of several silent *loci*.

(A) Expression of H3.1- (*HTR1* and *HTR9*) and H3.3 (*HTR5* and *HTR8*)-encoding genes analyzed by qRT-PCR in 2.5-week-old *in vitro*-grown WT and *atrx-1* plants. WT was set to 1.

(B) Ratio of histone H3K9me2/H3 occupancy at heterochromatin *loci* (TSI, 106B, 180bp, *Mule)* and at a subtelomeric region (At5g67640) assessed by ChIP qPCR in 2.5-week-old *in vitro*-grown WT, *atrx-1* and *hira-1* mutant plants.

(C) Expression of three heterochromatin repeats (TSI, 106B, 180bp) analyzed by qRT-PCR in 2.5-week-old *in vitro*-grown WT, *fas1-4* and *atrx-1* plants. The *fas1-4* mutant was used as a control for TSI silencing release (Pecinka et al., 2010). Student test, *, $p<0.005$. WT was set to 1.

(D) Analysis of *Mule* (At2g15810) expression in *atrx-1* mutants by semi-quantitative RT-PCR on three biological replicates. *At2g28390* was used as a control.

(E) Expression of three active genes (*UBC28, UEV1C, HXK1*) analyzed by qRT-PCR in 2.5-week-old *in vitro*-grown WT and *atrx-1* plants. WT was set to 1.

Supplemental Figure S5. Effects of ATRX loss on H3.1 levels.

(A) eH3.1 expression measured by qRT-PCR in eH3.1 lines with or without the *atrx-1* mutation. Two lines were analyzed for each genotype. Line 1 from eH3.1 was set to 1.

(B) Histone H3 occupancy at heterochromatic repeats (TSI, 106B, 180bp), at 5S rDNA *loci* and at three active genes (*UBC28, UEV1C, HXK1*) assessed by H3-ChIP qPCR in 2.5-week-old *in vitro*-grown eH3.1 and *atrx-1* eH3.1 plants. Student test, *, $p<0.05$.

Supplemental Figure S6. Effects of ATRX loss at 45S *loci*.

(A) Quantification of the relative of 45S rRNA variant abundance in 2.5-week-old *in vitro*-grown WT, *atrx-1* and *atrx-2* plants presented in Figure 6C. Student test, ***, $p<0.001$.

(B) Quantification of the relative 45S rDNA variant abundance presented in Figure 6D for ratios of histone H3/Input (left panel) and H3K9me2/Input (right panel) from the H3-ChIP and the H3K9me2-ChIP performed on 2.5-week-old *in vitro*-grown *atrx-1* mutant plants. Student test, **, $p<0.01$.

(C) Histone H3 occupancy at 45S rDNA *loci* assessed by H3-ChIP qPCR in 2.5-week-old *in vitro*-grown eH3.1 and *atrx-1* eH3.1 plants.

(D) RT-PCR on cDNA prepared from 2.5-week-old *in vitro*-grown eH3.1 and *atrx-1* eH3.1 plants. Detection of pre-rRNA transcripts was performed using primers surrounding the indels in the external transcribed 45S spacer region on three biological replicates.

(E) Quantification of 45S rRNA variant abundance from RT-PCR on cDNA presented in (D). Student test, ***, $p<0.001$.

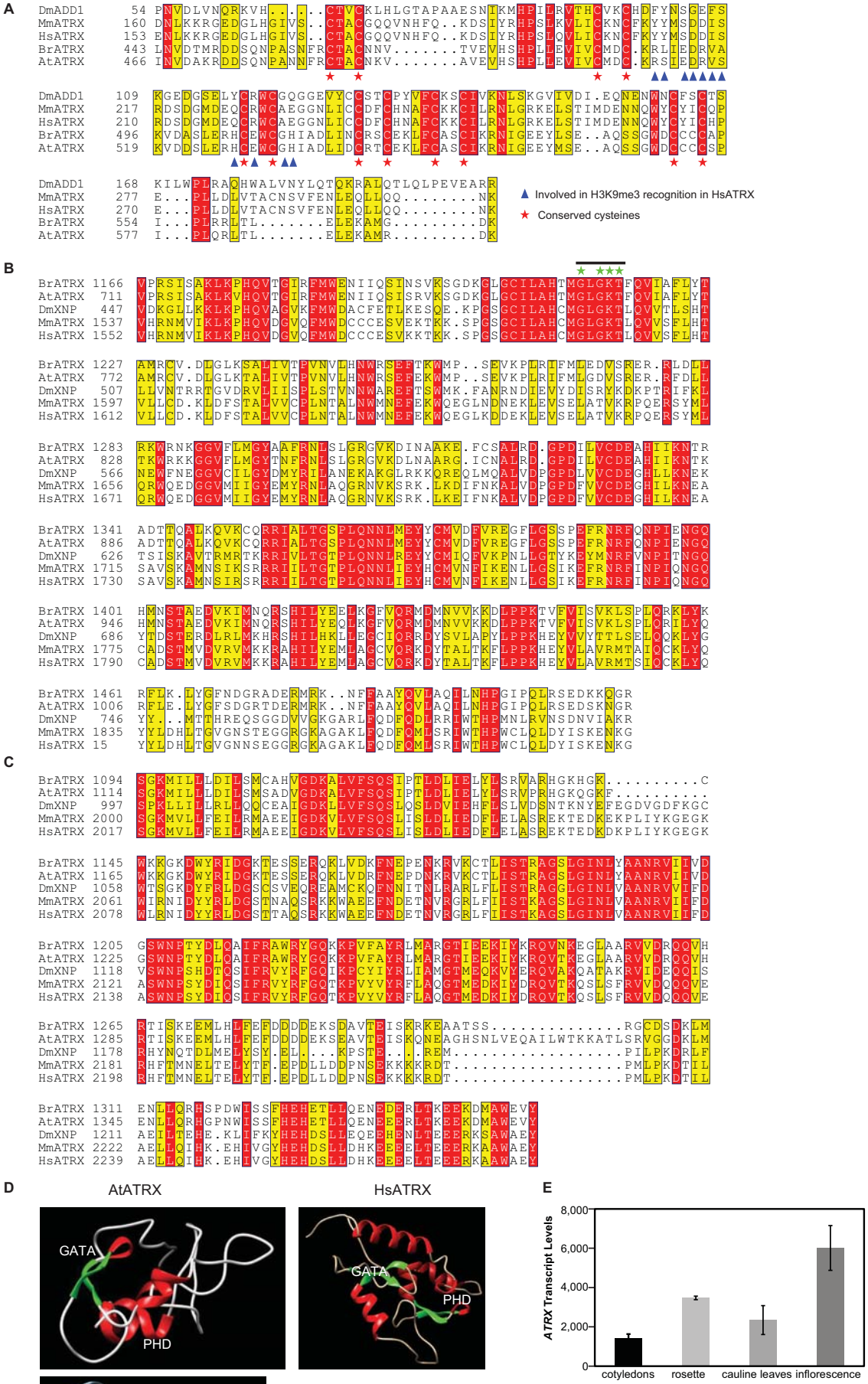
Supplemental Tables Legends

Supplemental Table S1. Segregation analysis in F2 populations from crosses between *atrx* alleles and mutants in the HIR complex.

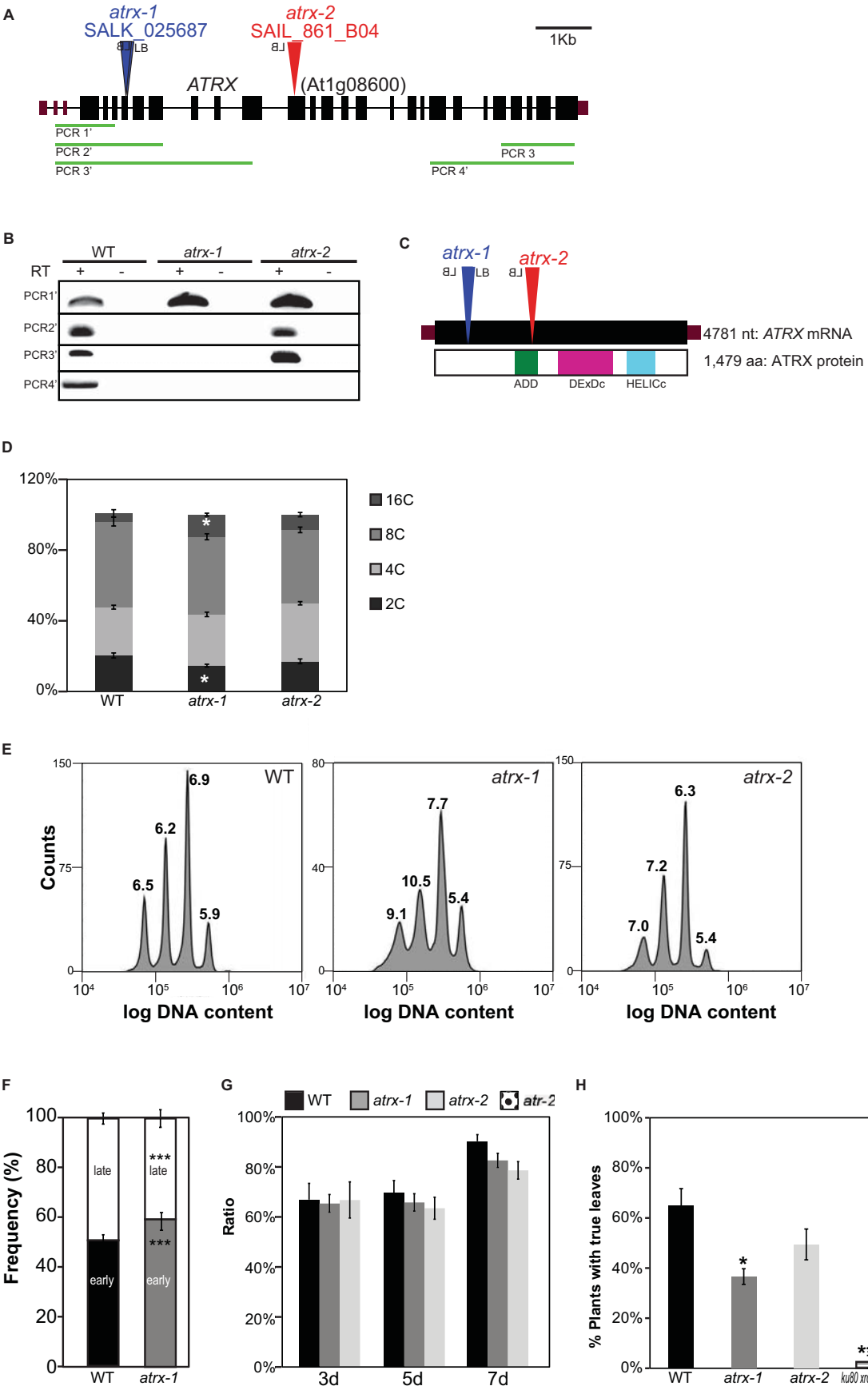
Supplemental Table S2. Segregation analysis in F3 populations from crosses between HIR complex mutants and *atrx* alleles.

Supplemental Table S3. Segregation analysis in F2 populations from crosses between *fas2-5* and *atrx* alleles.

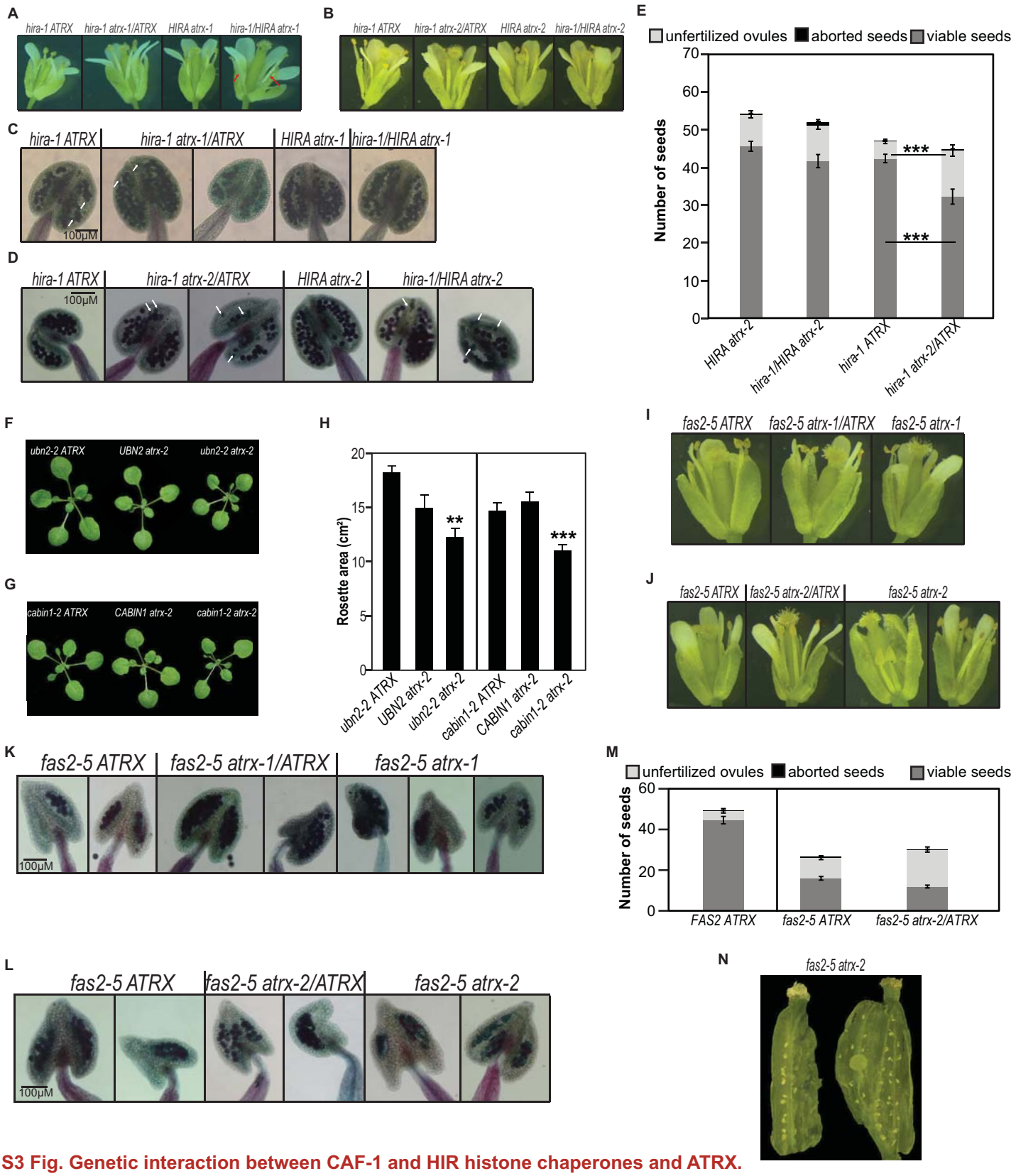
Supplemental Table S4. Primer list.



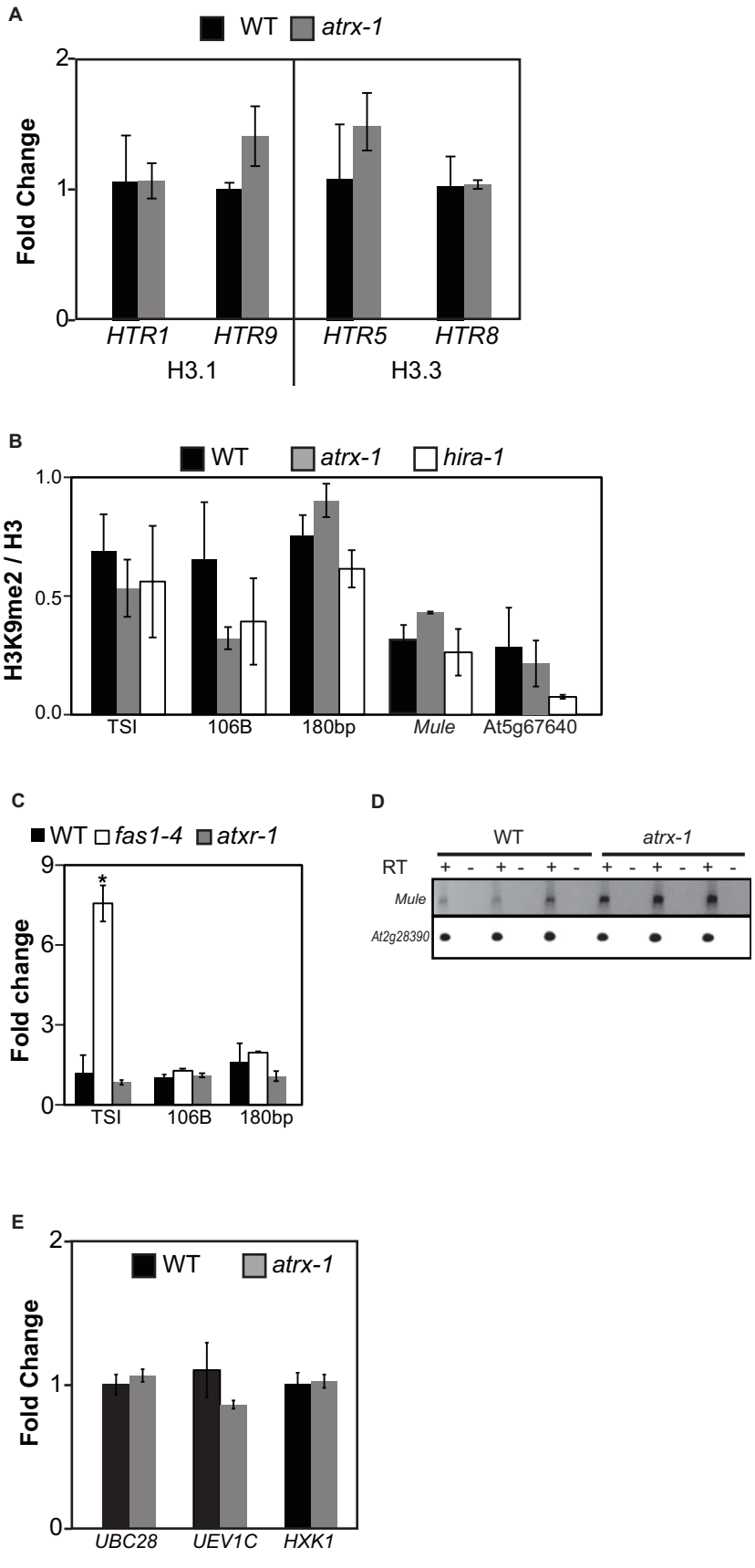
S1 Fig. Evolutionary conservation of the ATRX chaperones and expression of Arabidopsis *ATRX*.



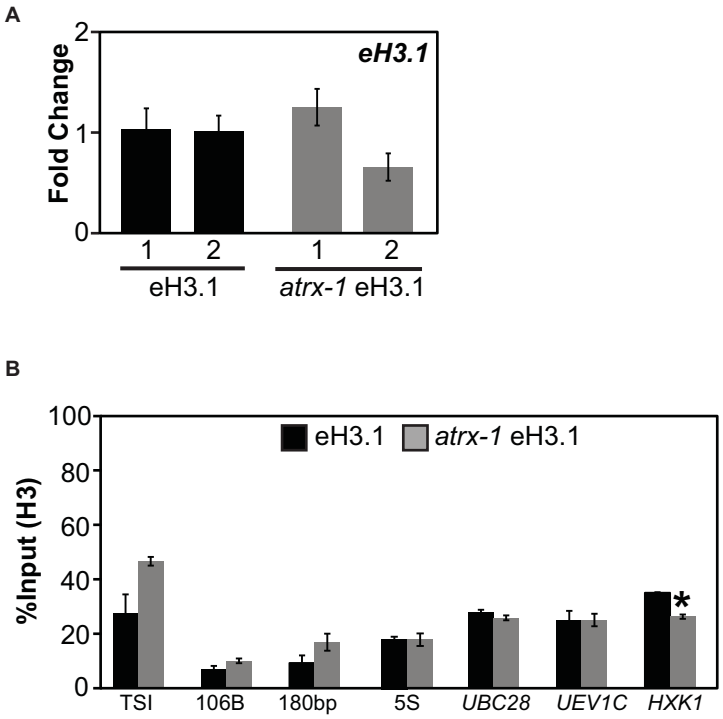
S2 Fig. Analysis of *atrx* mutant alleles.



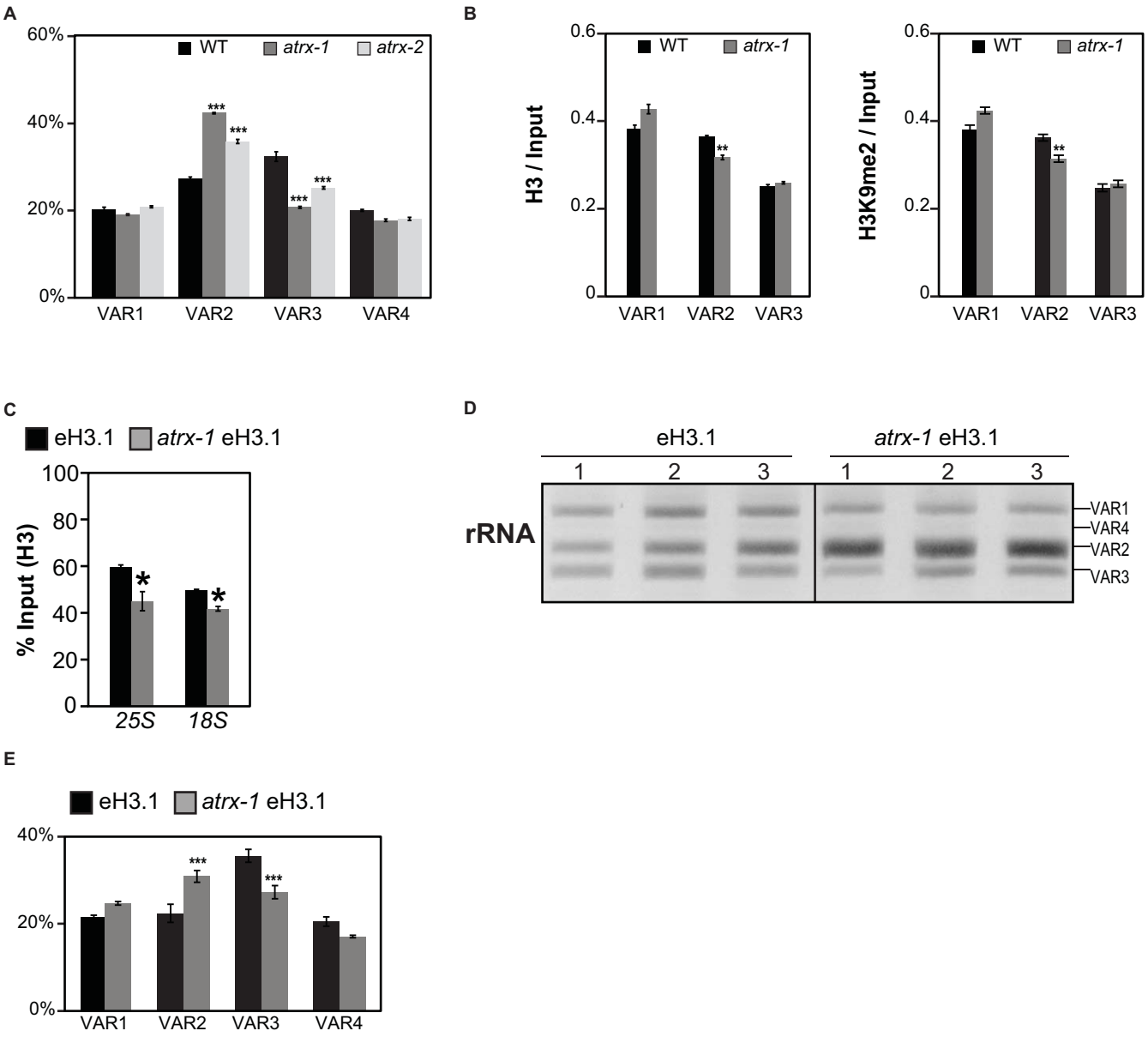
S3 Fig. Genetic interaction between CAF-1 and HIR histone chaperones and ATRX.



S4 Fig. Effects of ATRX loss on H3K9me2 enrichment and expression of several silent loci.



S5 Fig. Effects of ATRX loss on H3.1 levels.



S6 Fig. Effects of ATRX loss at 45S loci.

S1 Table. Segregation analysis in F2 populations from crosses between *atrx* alleles and mutants in the HIR complex.

A. Number and percentage of F2 plants with the indicated genotype. 111 F2 plants from one individual F1 plant were genotyped for *hira-1* and *atrx-1* mutations. Transmission of the *hira-1* mutation in the F2 population is affected (29 *HIRA*, 69 *hira-1/HIRA*, 13 *hira-1*): $\chi 2$ (2, N=111) = 10.1, P<0.01. n, number of plants obtained for the indicated genotype.

Genotype	n	Observed %	Expected %
HIRA atrx-1	5	4,5	6,94
hira-1 ATRX	6	5,4	6,94
HIRA ATRX	8	7,2	6,94
hira-1 atrx-1	1	0,9	6,94
hira-1/HIRA atrx-1	17	15,3	13,88
hira-1/HIRA ATRX	16	14,4	13,88
hira-1 atrx-1/ATRX	6	5,4	13,88
HIRA atrx-1/ATRX	16	14,4	13,88
hira-1/HIRA atrx-1/ATRX	36	32,4	27,75

$\chi 2$ (8, N= 111) = 14.2; not statistically significant.

B. Number and percentage of F2 plants with the indicated genotype. 115 F2 plants from one individual F1 plant were genotyped for *hira-1* and *atrx-2* mutations. Transmission of the *hira-1* mutation in the F2 population is affected (40 *HIRA*, 59 *hira-1/HIRA*, 16 *hira-1*): $\chi 2$ (2, N=115) = 10.8, P<0.01. n, number of plants obtained for the indicated genotype.

Genotype	n	Observed %	Expected %
HIRA atrx-2	9	7,8	7,19
hira-1 ATRX	5	4,3	7,19
HIRA ATRX	9	7,8	7,19
hira-1 atrx-2	0	0	7,19
hira-1/HIRA atrx-2	14	12,2	14,38
hira-1/HIRA ATRX	15	13,0	14,38
hira-1 atrx-2/ATRX	11	9,6	14,38
HIRA atrx-2/ATRX	22	19,1	14,38
hira-1/HIRA atrx-2/ATRX	30	26,1	28,75

$\chi 2$ (8, N= 115) = 13.7; not statistically significant.

C. Number and percentage of F2 plants with the indicated genotype. 102 F2 plants from one individual F1 plant were genotyped for *ubn2-2* and *atrx-1* mutations. n, number of plants obtained for the indicated genotype.

Genotype	n	Observed %	Expected %
UBN2 atrx-1	6	5,9	6,38
ubn2-2 ATRX	4	3,9	6,38
UBN2 ATRX	7	6,9	6,38
ubn2-2 atrx-1	5	4,9	6,38
ubn2-2/UBN2 atrx-1	7	6,9	12,75
ubn2-2/UBN2 ATRX	16	15,7	12,75
ubn2-2 atrx-1/ATRX	7	6,9	12,75
UBN2 atrx-1/ATRX	19	18,6	12,75
ubn2-2/UBN2 atrx-1/ATRX	31	30,4	25,50

$\chi 2$ (8, N= 102) = 11.5; not statistically significant.

D. Number and percentage of F2 plants with the indicated genotype. 101 F2 plants from one individual F1 plant were genotyped for *ubn2-2* and *atrx-2* mutations. n, number of plants obtained for the indicated genotype.

Genotype	n	Observed %	Expected %
UBN2 atrx-2	9	8,9	6,31
ubn2-2 ATRX	2	2,0	6,31
UBN2 ATRX	4	4,0	6,31
ubn2-2 atrx-2	6	5,9	6,31
ubn2-2/UBN2 atrx-2	6	5,9	12,63
ubn2-2/UBN2 ATRX	16	15,8	12,63
ubn2-2 atrx-2/ATRX	13	12,9	12,63
UBN2 atrx-2/ATRX	19	18,8	12,63
ubn2-2/UBN2 atrx-2/ATRX	26	25,7	23,25

$\chi 2$ (8, N= 101) = 12.6; not statistically significant.

E. Number and percentage of F2 plants with the indicated genotype. 101 F2 plants from one individual F1 plant were genotyped for *cabin1-2* and *atrx-1* mutations. n, number of plants obtained for the indicated genotype.

Genotype	n	Observed %	Expected %
CABIN atrx-1	9	8,9	6,31
cabin1-2 ATRX	7	6,9	6,31
CABIN1 ATRX	7	6,9	6,31
cabin1-2 atrx-1	6	5,9	6,31
cabin1-2/CABIN1 atrx-1	16	15,8	12,63
cabin1-2/CABIN1 ATRX	11	10,9	12,63
cabin1-2 atrx-1/ATRX	16	15,8	12,63
CABIN1 atrx-1/ATRX	9	8,9	12,63
cabin1-2/CABIN1 atrx-1/ATRX	20	19,8	25,25

$\chi 2$ (8, N= 101) = 5.5; not statistically significant.

F. Number and percentage of F2 plants with the indicated genotype. 101 F2 plants from one individual F1 plant were genotyped for *cabin1-2* and *atrx-2* mutations. n, number of plants obtained for the indicated genotype.

Genotype	n	Observed %	Expected %
CABIN atrx-2	9	8,9	6,31
cabin1-2 ATRX	6	5,9	6,31
CABIN1 ATRX	7	6,9	6,31
cabin1-2 atrx-2	2	2,0	6,31
cabin1-2/CABIN1 atrx-2	13	12,9	12,63
cabin1-2/CABIN1 ATRX	19	18,8	12,63
cabin1-2 atrx-2/ATRX	10	9,9	12,63
CABIN1 atrx-2/ATRX	12	11,9	12,63
cabin1-2/CABIN1 atrx-2/ATRX	23	22,8	25,25

$\chi 2$ (8, N= 101) = 8.2; not statistically significant.

S2 Table. Segregation analysis in F3 populations from crosses between HIR complex mutants and *atrx* alleles.

A. Number and percentage of F3 plants with the indicated genotype. F3 plants from several individual F2 plants coming from crosses between *hira-1* and *atrx-1* were genotyped for each mutation. n, number of plants obtained for the indicated genotype; N, number of independent F3 plants used in this study.

F2 Genotype	F3 Genotype	n	Observed %	Expected %	N
hira-1 atrx-1/ATRX	*ATRX*	10	45.5	25	1
	atrx-1/ATRX	12	54.5	50	
	atrx-1	0	0	25	
$\chi^2(2)= 9.3$; $P<0.01$					
hira-1/HIRA atrx-1	*HIRA*	15	35,7	25	2
	hira-1/HIRA	26	61,9	50	
	hira-1	1	2.4	25	
$\chi^2(2)= 11.7$; $P<0.005$					

B. Number and percentage of F3 plants with the indicated genotype. F3 plants from several individual F2 plants coming from crosses between *hira-1* and *atrx-2* were genotyped for each mutation. n, number of plants obtained for the indicated genotype; N, number of independent F3 plants used in this study.

F2 Genotype	F3 Genotype	n	Observed %	Expected %	N
hira-1 atrx-2/ ATRX	*ATRX*	30	58	25	3
	atrx-2/ATRX	22	42	50	
	atrx-2	0	0	25	
$\chi^2(2)= 35.8$; $P<0.005$					
hira-1/HIRA atrx-2	*HIRA*	54	56,8	25	5
	hira-1/HIRA	41	43,2	50	
	hira-1	0	0	25	
$\chi^2(2)= 63.2$; $P<0.005$					

C. Number and percentage of F3 plants with the indicated genotype. F3 plants from several individual F2 plants coming from crosses between *cabin1-2* and *atrx-1* were genotyped for each mutation. n, number of plants obtained for the indicated genotype; N, number of independent F3 plants used in this study.

F2 Genotype	F3 Genotype	n	Observed %	Expected %	N
cabin1-2 atrx-1/ATRX	*ATRX*	11	27,5	25	2
	atrx-2/ATRX	24	60,0	50	
	atrx-2	5	12,5	25	
$\chi^2(2)= 3.5$; not statistically significant					
cabin1-2/CABIN1 atrx-1	*CABIN1*	10	23,3	25	2
	cabin1-2/ CABIN1	27	62,8	50	
	cabin1-2	6	14,0	25	
$\chi^2(2)= 3.6$; not statistically significant					

D. Number and percentage of F3 plants with the indicated genotype. F3 plants from several individual F2 plants coming from crosses between *cabin1-2* and *atrx-2* were genotyped for each mutation. n, number of plants obtained for the indicated genotype; N, number of independent F3 plants used in this study.

F2 Genotype	F3 Genotype	n	Observed %	Expected %	N
cabin1-2 atrx-2/ATRX	*ATRX*	31	36,0	25	3
	atrx-2/ATRX	43	50,0	50	
	atrx-2	12	14,0	25	
$\chi^2(2)= 8.4$; $P<0.025$					
cabin1-2/CABIN1 atrx-2	*CABIN1*	25	38,5	25	2
	cabin1-2/ CABIN1	30	46,2	50	
	cabin1-2	10	15,4	25	
$\chi^2(2)= 7.3$; $P<0.05$					

S3 Table. Segregation analysis in F2 populations from crosses between *fas2-5* and *atrx* alleles.

A. Number and percentage of F2 plants with the indicated genotype. 77 F2 plants from one individual F1 plant were genotyped for *fas2-5* and *atrx-1* mutations. n, number of plants obtained for the indicated genotype.

Genotype	n	Observed %	Expected %
FAS2 atrx-1	5	6,5	4,81
fas2-5 ATRX	4	5,2	4,81
FAS2 ATRX	6	7,8	4,81
fas2-5 atrx-1	4	5,2	4,81
fas2-5/FAS2 atrx-1	10	13,0	9,63
fas2-5/FAS2 ATRX	10	13,0	9,63
fas2-5 atrx-1/ATRX	9	11,7	9,63
FAS2 atrx-1/ATRX	10	13,0	9,63
fas2-5/FAS2 atrx-1/ATRX	19	24,7	19,25
$\chi 2$ (8, N= 77) = 0.7; not significant			

B. Number and percentage of F2 plants with the indicated genotype. 69 F2 plants from one individual F1 plant were genotyped for *fas2-5* and *atrx-2* mutations. n, number of plants obtained for the indicated genotype.

Genotype	n	Observed %	Expected %
FAS2 atrx-2	4	5,8	4,31
fas2-5 ATRX	6	8,7	4,31
FAS2 ATRX	5	7,2	4,31
fas2-5 atrx-2	6	8,7	4,31
fas2-5/FAS2 atrx-2	5	7,2	8,63
fas2-5/FAS2 ATRX	11	15,9	8,63
fas2-5 atrx-2/ATRX	7	10,1	8,63
FAS2 atrx-2/ATRX	13	18,8	8,63
fas2-5/FAS2 atrx-2/ATRX	12	17,4	17,25
$\chi 2$ (8, N= 69) = 7.8; not significant			

S4 Table. Primer list.

Gene	Purpose	FORWARD (F) and REVERSE (R) PRIMERS (5' to 3')
ATRX	*atrx-1* genotyping	ATRX_Salk025687_LP2 : TCTTCTGGCAGTTGAGAGCA ATRX_Salk025687_RP : GTCAAGCTCAGATGTTCCAGC LBb1.3: ATTTTGCCGATTTCGGAAC
ATRX	*atrx-2* genotyping	LBR Sail: TAGCATCTGAATTTCATAACCAATCTCGATACAC ATRX_Sail861B04_LP: AGGAACCCTCACAGCTTCTTC ATRX_Sail861B04_RP:TCACATGGATGGCTTCTTTTC
ATRX	Analysis of ATRX transcripts	PCR1: GCTGCAGAAGCTCAAGAAG and GTCAAGCTCAGATGTTCCAGC PCR2: TTATTAGGTGTGCGGCAGAG and CTTCCGCACCTTCTGGAATA PCR3: AAGGTGTCGGACTTCAGTGG and GGGACCCGTTGAACTTCCTCCC PCR1': GATATATGCCGCAAGGATGC and CTTCTTGAGCTTCTGCAGC PCR2': GATATATGCCGCAAGGATGC and GTCAAGCTCAGATGTTCCAGC PCR3': GATATATGCCGCAAGGATGC and CTCTGCCGCACACCTAATAA PCR4': GAATTCAACTGCGGAGGATG and GGGACCCGTTGAACTTCCTCCC
ATRX	qRT-PCR	ATRX_RT_For: AGCAATCTGGTGGAGCAAGCAA ATRX_RT_Rev: GGGACCCGTTGAACTTCCTCCC
25S	ChIP q-PCR	25S f: GCATCAGGTCTCCAAGGTG 25S r: AGCCCTCAGAGCCAATC
18S	ChIP q-PCR	18S f: CGTAGTTGAACCTTGGGATG 18S r: CACGACCCGGCCAATTA
45S	Analysis of 45S variants	5allrRNAvar: GACAGACTTGTCCAAAACGCCCACC 3allrRNAvar: CTGGTCGAGGAATCCTGGACGATT
RAD51 (At5g20850)	qRT-PCR	CTCCGAGGAAGGATCTCTTGCAG GCTCGCACTAGTGAACCCCAGAGG
BRCA1 (At4g21070)	qRT-PCR	GTTACGTGTGCAAAACTCATACCAGAATG GATACTTGTTTAGGCTGAGAGTGCAGTGG
UBC28 (At1g64230)	qRT-PCR	TCCAGAAGGATCCTCCAACTTCCTGCAGT ATGGTTACGAGAAAGACACCGCCTGAATA

PARP2 (At4g02390)	qRT-PCR	ACATGGTTTACACCAGATGGGGAAGAG GGACTTGGGATGTGGGATAAACTCCTT
HXK1 (At4g29130)	ChIP qPCR	At4g29130_ChIP-F: AGGAGCTCGTCTCTCTGCTG At4g29130_ChIP-R: GCTCAAACAATCCACCATCC
UEV1C (At2g36060)	ChIP qPCR	At2G36060_ChIP-F: GGTGACTGAAATGTGAATTTGC At2G36060_ChIP-R: ATGCAGCCATCTCCTTCTTC
UBC28 (At1g64230)		At1g64230-ChIP-F: TCATTGTTAACGGACCCAAAC
		At1g64230-ChIP-R: CCAGCTTCTCGCAGTAGACTC
TSI	ChIP qPCR and qRT-PCR	TSIq-F: CTCTACCCTTTGCATTCATGAATCCTT TSIq-R: GATGGGCAAAAGCCCTCGGTTTTAAAATG
At2g28390	qRT-PCR	SA-F: AACTCTATGCAGCATTTGATCCACT SA-R: TGATTGCATATCTTTATCGCCATC
Ta3	semi quantitative RT-PCR	Ta3 middle-F : GATTCTTACTGTAAAGAACATGGCATTG Ta3 middle-R : TCCAAATTTCCTGAGGTGCTTGTAACC
Mule (At2g15810)	semi quantitative RT-PCR and ChIP q-PCR	MULE-F2 : CTGTCCGCGAGTGTCATCAAGTAGC MULE-R2 : GATACTTGTTGACAAGTGTTTAGCAAGCC
106B	qRT-PCR and ChIP qPCR	AP394 106Bq-F: TCATTATGCTAGGTGGTTGA AP395 106Bq-R: GACAACAAGTTCATTAACCA
180bp	qRT-PCR and ChIP qPCR	180(all)-F: ACCATCAAAGCCTTGAGAAGCA 180(all)-R: CCGTATGAGTCTTTGTCTTTGTATCTTCT
HTR1 (At5g65360)	qRT-PCR	At5g65360_HTR1_qpcr_Forbis: AGCGATCTCACGAACCAAAC At5g65360_HTR1_qpcr_Revbis: GAAATCCACCGGAGGAAAAG
HTR8 (At5g10980)	qRT-PCR	At5g10980_HTR8_qpcr_Forbis: AGCGATCTCACGAACCAAAC At5g10980_HTR8_qpcr_Revbis: GAAATCCACCGGAGGAAAAG
HTR5 (At4g40040)	qRT-PCR	At4g40040 RT-F: AAT GCC CAA AGA CAT TCA GC At4g40040 RT-R: CAT TGC CAA AGA AGA AAG CA
HTR9 (At5g10400)	qRT-PCR	HTR9_RT-F: TAATCTCTGTGCGATTCATGCT HTR9_RT-R: CGAAAACGAAAAGAGACAGCTT
Intergenic region	ChIP qPCR	IG-2g17670-80qF: GGCTACTGTCTAGTTCATATCTTAGA

between At2g17670 and At2g17680		IG-2g17670-80qR: TAGGTTGGCATCCGATCCAGAGT
eH3.1	qRT-PCR	HTR9_RT-F: TAATCTCTGTGCGATTCATGCT tag_oligo_rt : CACGTCGTAGGGGTATCCTC
At1g01240	ChIP qPCR	At1g01240_For: GGTGCTGGTTTGCTCTTAGG At1g01240_Rev: ACTCTTGAAAAGATGAGTGACC
At3g63180	ChIP qPCR	At3g63180_For :TAACTTCGTGCCCTGGTCATC At3g63180_Rev: ACCGATCCAAAGCGTTCACTC
At5g67640	ChIP qPCR	At5g67640_For :TGGTAGGACTAATCGGTGAC At5g67640_Rev:GGAATTTCAGGGATTGCATTGG
Athila LTR pericentromeric retrotransposon	semi quantitative RT-PCR	Athila LTR-F :TGTTTCATCCACGTTTCATCTC Athila LTR-R :AGCAATAAGCGCAACTAATCC
Mule transposon At1g40097	semi quantitative RT-PCR	Mule At1g40097-F :GGTTTTGATACCGAATTTTG Mule At1g40097-R :AGCGGAGGAATATACAACTC
Mule transposon At1g43280	semi quantitative RT-PCR	Mule At1g43280-F :GGTTAGGAAAGTGAAGCTTGAG Mule At1g43280-R :CCAGTGAGACAAAGGCATAC
Mu1 (At4g08680)	semi quantitative RT-PCR	AtMU1-R: CTTAGCCTTCTTTTCAATCTCA AtMU1-F: GTGGATATACCAAAAACACAA
CACTA1 transposon	semi quantitative RT-PCR	SP15: AACAAAAGCATCATTCTACTTAAC SP40: AGGCCTACAATGGAAATGACG
AtSN1 (At3TE63860)	semi quantitative RT-PCR	AtSN1_qpcr_For: AAAGAAGATGAATTTCTGGTATGG AtSN1_qpcr_Rev: AGCCTAGTTTTAATTCTACGGATCA
T5L23.26 CACTA-like transposon	semi quantitative RT-PCR	T5L23.26-F:GGCTAGCTGTCCGACTCAATGACCT T5L23.26-R:CAGACATCCTTTCCTTCAGCTTAGC
COPIA78 LTR transposon	semi quantitative RT-PCR	COPIA78qF2:CGGTGCTCACAAAGAGCAACTATG COPIA78qR3:ATCCTTGATAGATTAGACAGAGAGCT
5S	ChIP qPCR	RTPCR5S1:GGATGCGATCATACCAG 5SUNIV2:CGAAAAGGTATCACATGCC

5S	qRT-PCR	RTPCR5S1:GGATGCGATCATACCAG 5S_2R:GGGAGGTCACCCATCCTAGT