



HAL
open science

Etude des gènes METHYLTRANSFERASE 1 chez *Triticum aestivum* : identification, histoire évolutive et création de lignées RNAi

Mélanie Thomas

► **To cite this version:**

Mélanie Thomas. Etude des gènes METHYLTRANSFERASE 1 chez *Triticum aestivum* : identification, histoire évolutive et création de lignées RNAi. Sciences agricoles. Université Blaise Pascal - Clermont-Ferrand II, 2014. Français. NNT : 2014CLF22478 . tel-01126978

HAL Id: tel-01126978

<https://theses.hal.science/tel-01126978>

Submitted on 6 Mar 2015

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

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

**ECOLE DOCTORALE
DES SCIENCES DE LA VIE, SANTE, AGRONOMIE, ENVIRONNEMENT
N° d'ordre 635**

Thèse

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

DOCTEUR D'UNIVERSITE
Spécialité : Physiologie et Génétique Moléculaires

Soutenue le 10 juillet 2014

Mélanie THOMAS

Etude des gènes *METHYLTRANSFERASE 1* chez
Triticum aestivum : identification, histoire évolutive et
création de lignées RNAi

Président :
Rapporteurs : Dr. GUIDERDONI Emmanuel, CIRAD Montpellier
Pr. MAURY Stéphane, Université d'Orléans
Dr. LEBLANC Olivier, IRD Montpellier
Examineur : Dr. PRAUD Sébastien, BIOGEMMA Clermont-Ferrand
Directeur de thèse : Pr. TATOUT Christophe, GReD Aubière

Laboratoire UMR GReD CNRS-INSERM-Clermont Université
24 avenue des Landais 63171 Aubière

BIOGEMMA
Site de la Garenne, Route d'Ennezat 63720 Chappes.



Voici venu le temps d'écrire les dernières lignes de ce rapport et de remercier toutes les personnes qui m'ont aidé et soutenue ; de remercier la vie/le hasard/ le destin de m'avoir fait croiser la route de certaines d'entre elles pour une durée plus ou moins longue.

Je ne ferai pas dans l'originalité pour le début de chaque paragraphe et me contenterai d'un « Merci » (parfois agrémenté d'un « Grand »).

Le premier de mes remerciements va tout naturellement à mon directeur de thèse, Christophe. Merci de m'avoir pris en stage en M1. Merci pour ce mail que j'ai reçu alors que j'étais à l'autre bout du monde. Merci d'avoir pensé à moi et de m'avoir proposé ce sujet. Merci d'avoir toujours été là pour répondre à mes questions, m'aider à organiser mes idées et à faire le point. Merci de m'avoir permis de vivre cette « folle » aventure.

Merci à Jacques d'avoir été le lien avec Biogemma. Merci pour ta gentillesse et de t'être toujours rendu disponible quand j'en ai eu besoin. Merci d'avoir toujours été positif, une vraie bouffée d'air pendant les périodes de doutes !

Merci à Wyatt et Christophe S. d'avoir fait une place au sein de l'équipe GMTD pour une thématique comme celle qui fut la mienne.

Merci à Sylvette d'avoir accepté de faire une place pour le blé au sein des fondamentalistes d'arabido !

Merci à Aline pour tes conseils scientifiques toujours pertinents, judicieux et éclairants. Merci pour le temps que tu m'as consacré et pour l'exemple.

Merci à l'ensemble de l'équipe GMTD et en particulier Christian, David et Ayhan pour leur aide dans le premier mois de la thèse et pour avoir facilité mes débuts.

Merci à Jean Philippe, Alexandra, Jorge et Hervé pour leur implication dans mon projet, et leur patience : ça ne va jamais aussi vite et aussi bien qu'on le voudrait.

Merci à Arienne pour le temps passé à me guider sur ces machines qui m'étaient inconnues et à me former sur les manips. Merci à Arienne et Nadine pour leur gentillesse et leur bonne humeur constante et communicative.

un merci à Mickael pour toutes ces analyses bioinfos, pour le temps passé à expliquer toutes ces notions qui m'étaient inconnues.

Merci à Sébastien F pour les conseils, les remarques pertinentes, le calme constant et la positive attitude !

Merci à l'ensemble des personnes de l'équipe CIV, de l'équipe analyse et de l'équipe serre pour tous les soins et le suivi de mes petites plantes !!

Merci à Etienne d'avoir réussi à nous caser aussi souvent dans ton emploi du temps déjà bien rempli. Un grand merci de m'avoir rassuré lors de mon premier congrès !

un très grand merci à Lise pour les analyses vol1 et vol2! Merci d'avoir pris autant de temps sur ton propre projet de thèse.

Merci à Frederic Choulet et Pierre Sourdille pour la relecture du papier et pour les graines.

Merci à Caroline Tassy et Pierre Barret d'avoir été nos « sauveurs ! », pour la place et l'aide en serre.

Merci à Clémentine et Philippe d'avoir participé à mes comités de thèse.

Merci à Emmanuel Guiderdoni, Olivier Leblanc et Stéphane Maury d'avoir accepté d'être membre de mon jury.

Merci à Sébastien P d'avoir été un président de jury aussi cool !

Merci à l'ensemble des personnes qui ont répondu à une de mes questions, indiqué le lieu de rangement d'un produit, montré le fonctionnement d'une machine...

Merci à Marie Claude, le pilier du labo. Merci pour les boîtes, tes tiroirs foisonnant de trésors. Un grand merci pour toutes ces discussions, pauses thé-café, coups de gueule et coup de cœur. Je suis bien contente d'être partie avant toi, le labo m'aurait paru bien vu vide sinon.

Merci à Sylvie et Sylvianne, également connues comme Tic et Tac, pour les corrections orthographiques (loin d'être mon fort !). Merci pour les pauses repas et pauses thé, les solutions que je vous ai piqué (avec votre accord).

Merci à Sophie et Sylvie pour l'aide au labo sur les derniers mois.

Merci à Sam d'avoir imposé France Inter dans le labo, je n'écoute plus qu'elle maintenant. Merci pour toutes ces fois où je suis venu te déranger pour tout et n'importe quoi.

Merci à Manu pour toutes ces discussions plus ou moins animés. Étonnamment, il nous arrive assez souvent d'être d'accord.

Merci à Céline pour ta rigueur et ta disponibilité au travail, ton ouverture d'esprit. Je te souhaite d'être condamnée à rester au milieu de ces fous !!

Merci à Elodie pour les pauses cafés, les infos croustillantes et pour tous les « Michel ».

Merci à Margot, Heidi, Angeles, Letitia, Sébastien, Pierre, Emilie pour toutes les pauses au soleil ou dans la cuisine, souvent avec cookies ! les dej' collectifs pizza ou bagels (bien meilleurs), les séances de pestage contre les chefs et autres membres du labo...

Merci à Marie-No de m'avoir fait une place dans son bureau.

Merci à Laurianne, Axel et Matthias. Je pensais faire des remerciements séparés pour chacun d'entre vous mais je renonce de peur de finir trop larmoyante ! Je vous remercie pour tous ces moments passés ensemble au labo et en dehors. Je vous remercie d'avoir toujours été d'un soutien sans faille : thésard toujours solidaires ! Dans le désordre : merci pour les cris dans les couloirs, les sourires et les questions gentilles en réunion d'équipe, les petits poneys, les messes basses, les fous rires, les prêts de livres, les repas à la maison, les répétitions de présentations entre nous, les blagues de plus ou moins bon gout..... pour être exhaustive, il me faudrait reprendre presque jour par jour ces 3 ans et demi de thèse. Donc merci pour tout, merci d'avoir été plus que des collègues, des amis.

Merci à mon père de m'avoir soutenu dans ces longues études, d'avoir toujours cru en moi et de ne m'avoir jamais mis de frein.

Merci à Marion d'être ma sœur ! Merci d'être toujours dispo pour une conversation au téléphone (quelque soit sa durée), d'être toujours prête à débarquer à la maison si besoin. Merci d'être fière de moi.

Merci à ma famille et mes amis de vous être déplacé à ma soutenance et pour les messages de soutien. Et merci d'avoir compris mon manque de disponibilité, la fatigue et l'omniprésence du sujet « thèse » ces dernières années.

Un infini merci à Romain, mon amour. Merci de m'avoir suivi à Clermont, merci pour ton soutien inconditionnel, d'avoir fait parfois plus que ta part à la maison. Merci de m'aimer si fort.

Merci à Armand, seule personne pour qui, thèse ou pas, ça ne changeait rien. C'est fou ce qu'on peut être efficace quand on veut rentrer avant l'heure du bain !!

Le blé tendre ou *Triticum aestivum* possède un génome hexaploïde ($2n=6X=42$ chromosomes) de très grande taille (17 Gb) formé de trois génomes diploïdes homéologues. Afin de répondre aux nouvelles contraintes sociétales et environnementales, de nouvelles variétés de blé doivent être créées. L'amélioration des variétés peut se baser sur la variabilité génétique, mais également sur les modifications épigénétiques mises en évidence ces dernières années. La méthylation de l'ADN est l'une de ces modifications, et se retrouve sous forme de 5-méthylcytosine (5mC). Le gène *METHYLTRANSFERASE1* (*MET1*) est bien connu pour son rôle dans le maintien de la méthylation de l'ADN au niveau des 5mC en contexte CpG chez *Arabidopsis*.

Afin d'identifier les orthologues de *MET1* chez le blé, la méthode de capture de séquence génomiques sur puce à ADN a été combinée avec des analyses bioinformatiques réalisées sur les récentes données de séquençage du génome du blé. J'ai identifié neuf copies du gène *TaMET-1* sur les chromosomes 2, 5 et 7, et ce pour les trois génomes homéologues. Deux événements de duplications géniques semblent être à l'origine de ces neuf copies. Suite à la seconde duplication, les copies du chromosome groupe 5 (groupe 5) ont évolué plus rapidement pour devenir des pseudogènes, peu ou pas exprimés. L'analyse de données d'expression par RNA-Seq a révélé que les copies des groupes 2 sont 10 à 40 fois plus exprimées que celles du groupe 7. Nous avons montré, pour les régions promotrices des groupes 5 et 7, la relation existant entre un faible niveau d'expression, une forte vitesse évolutive et un enrichissement en CpG, ce dernier étant associé avec une forte méthylation.

Plusieurs stratégies ont été envisagées afin de valider les fonctions des gènes *TaMET-1* : le crible de population de TILLING (Targeting Induced Local Lesions in Genomes), la construction de lignées RNAi ciblant *MET1* et l'utilisation de lignées délétées pour la partie du chromosome portant la copie *TaMET-1* la plus exprimée. Aucun mutant de TILLING n'a été identifié. Une analyse par conversion au bisulfite est actuellement en cours sur les lignées RNAi et lignées de délétion afin de valider un effet possible sur la méthylation. Ce travail permettra de conclure sur l'identification de lignées capables de modifier les profils de méthylation et qui pourraient être le point de départ pour induire de la variabilité épigénétique.

Mots clés : méthylation de l'ADN ; MET1 ; blé tendre ; lignées RNAi.

Bread wheat or *Triticum aestivum* possesses a large hexaploid genome ($2n=6X=42$ chromosomes, 17 Gb) formed by three homeologous diploid genomes. To better respond to new societal and environmental constraints, new wheat varieties have to be created. Breeding can use genetic variability and epigenetics modifications highlighted in recent years. DNA methylation is one of the epigenetic marks and is found as 5-methylcytosine (5mC). The *METHYLTRANSFERASE1* gene (*MET1*) is known to be involve in maintenance of 5mC DNA methylation in CpG context.

To identify *MET1* genes in wheat, a method based on genomic sequence capture and bioinformatics analyses on data from wheat genome were combined. I identified nine copies of *TaMET-1* gene on chromosomes 2, 5 and 7, for the three homeologous genomes. These nine copies seem to originate from two duplication events. After the second one, copies from chromosome 5 (group 5) evolved faster to become pseudogenes, not expressed or at a low level. Analysis of RNA-Seq expression data revealed that group 2 copies are expressed from 10 to 40 times more than the ones from group 7. For the promoter regions of group 5 and 7, we have shown a relationship between low expression level, high evolution rate and CpG enrichment, which is associated with high DNA methylation level.

Several strategies were chosen to validate *MET1* gene function : screen of TILLING population, construction of RNAi lines and use of deleted-chromosome line for the most expressed *TaMET-1* gene. No TILLING mutants were found. An analysis based on bisulfite conversion is in progress on RNAi lines and deleted lines to validate a putative effect on DNA methylation. This work will permit to identify lines able to modify DNA methylation pattern which will be the starting point to induce epigenetics variability.

Keywords : DNA methylation ; MET1 ; bread wheat ; RNAi lines.

Sommaire

Sommaire	1
Liste des abbréviations.....	7
Liste des illustrations.....	11
Synthèse bibliographique.....	15
Chapitre 1 : le blé	17
1. Données économiques.....	17
2. Histoire évolutive du blé tendre.....	19
3. Le génome du blé	21
3.1. Comparaison des trois génomes homéologues du blé tendre.....	21
3.2. Organisation générale du génome du blé	23
3.3. Le séquençage du génome du blé.....	25
Chapitre 2 : épigénétique & méthylation de l'ADN.....	29
1. Les différents niveaux de compaction de la chromatine.....	29
1.1. Euchromatine et hétérochromatine	31
1.2. Organisation de la chromatine en fonction de la taille des génomes	33
2. Chromatine et différents niveaux de régulation épigénétique.....	35
3. La méthylation de l'ADN	41
3.1. La méthylation chez les organismes vivants.....	41
3.1.1. Les méthylations de l'ADN.....	41
3.1.2. La méthylation de l'ADN chez les animaux et les végétaux.....	41
3.1.3. Déméthylation active de l'ADN	45
3.2. Rôles de la méthylation	47
3.2.1. Méthylation et augmentation de la mobilité des ETs.....	47
3.2.2. Méthylation et transcription.....	49
3.2.3. Méthylation et recombinaison.....	53
3.3. Mise en place et maintien des profils de méthylation chez les plantes	55
3.3.1. Contexte de méthylation, proportion et répartition	55
3.3.2. Mise en place et maintien des méthylations.....	59
3.3.2.1. <i>La voie RNA-directed DNA Methylation (RdDM)</i>	59
3.3.2.2. <i>Maintien de la méthylation en contexte asymétrique CpHpH</i>	61
3.3.2.3. <i>Maintien de la méthylation en contexte symétrique CpG et CpHpG</i>	61
3.3.2.4. <i>Implication du gène DECREASE IN DNA METHYLATION (DDM1) dans le</i> <i>maintien de la méthylation</i>	63
3.4. Le gène <i>METHYLTRANSFERASE 1</i>	63
3.4.1. Structure du gène MET1	63
3.4.2. MET1 et maintien des méthylations CpG	65
3.4.3. Action de MET1 pendant la réplication.....	67
3.4.4. MET1 et les marques histones	69
3.4.5. MET1 et la répression transcriptionnelle des éléments transposables.....	71
3.4.6. MET1 et le développement.....	71
Objectifs	75
Résultats	81
Chapitre 1 : Les gènes <i>TaMET-1</i> chez le blé tendre.....	83
1. Historique de la découverte des gènes <i>MET1</i> chez le blé tendre.....	83
2. Mise au point de la conversion au bisulfite	89
3. Préambule.....	91

4. Article : Genome dynamics at Methyltransferase 1 genomic regions in hexaploid wheat reveals the involvement of CG DNA methylation in genome evolution	93
5. Conclusion du chapitre 1	215
5.1. Histoire évolutive des gènes <i>MET1</i> chez les graminées	215
5.2. Vitesse évolution des gènes <i>MET1</i> :	217
Bilan	221
Chapitre 2 : étude des profils de méthylation.....	223
1. Lignées TILLING, RNAi, délétion	223
1.1. Mutants TILLING	223
1.1.1. Populations de TILLING	223
1.1.2. Stratégie de crible des populations	225
1.1.3. Résultats du crible des populations de mutagenèse	227
1.2. Lignées RNAi	229
1.2.1. Construction des vecteurs et clonage	231
1.2.2. Première vague de transformants RNAi <i>MET1</i>	231
1.2.3. Recherche de siRNA dirigés contre <i>MET1</i>	233
1.2.4. Analyses phénotypiques des lignées RNAi.....	233
1.2.5. Seconde vague de transformants RNAi <i>MET1</i>	235
1.3. Lignées de délétion.....	237
1.4. Bilan de l'obtention de lignées dérégulées pour <i>TaMET-1</i>	239
2. Etude des profils de méthylation	239
2.1. Profils de méthylation de gènes candidats	239
2.1.1. Choix des gènes candidats	241
2.1.1.1. Gènes du rendement	241
2.1.2. Choix des régions géniques étudiées.....	243
2.2. Séquençage MiSeq	245
2.2.1. Mise au point	245
2.2.1.1. Technique de séquençage : MiSeq	245
2.2.1.2. Stratégie d'amplification	247
2.2.1.3. Mise au point des amplifications	249
2.2.2. Bilan des génotypes étudiés.....	249
2.2.3. Production des amplicons pour séquençage.....	249
Discussion & perspectives.....	253
1. Quelles sont les principales conclusions de mon projet de thèse ?.....	255
1.1. Le gène <i>MET1</i> est très conservé au cours de l'évolution et présent chez le blé	255
1.2. Le blé compte neuf copies de <i>TaMET-1</i>	257
1.2.1. Elimination des paralogues après une Whole Genome Duplication (WGD)	257
1.2.2. Devenir des paralogues si ils ne sont pas éliminés.....	259
1.3. Les gènes <i>TaMET-1</i> du groupe 2 sont exprimés dans les tissus actifs pour la réplication.....	259
1.4. Il existerait une régulation transcriptionnelle de la famille <i>MET1</i> par la méthylation	261
1.5. Existe-t-il une autorégulation dans la famille <i>MET1</i> chez le blé ?.....	263
1.6. La perte de fonction de <i>MET1</i> conduirait à une létalité et/ou une stérilité	265
1.6.1. Constat suite à mon projet de thèse	265
1.6.2. Comment améliorer notre démarche ?	265
1.7. Questions futures	267
2. Mon projet de thèse aura-t-il des applications futures chez Biogemma ?.....	267
2.1. Méthode d'évaluation des lignées RNAi via l'efficacité de production de siRNA	267
2.2. Méthode d'étude du méthylome chez le blé par conversion au Bisulfite et séquençage	269
2.3. Modifier le méthylome du blé en ciblant <i>DDMI</i>	271
2.4. Effets possibles sur la dérégulation des éléments transposables dans les mutants <i>met1</i> et <i>ddm1</i>	273
2.5. Création d'EpiRIL <i>met1</i>	273
2.6. Utilisation de lignées RNAi pour changer les profils de recombinaison.....	275
3. Conclusion.....	277

Matériels & Méthodes.....	279
1. Identification des ESTs spécifique à <i>MET1</i> et <i>DDM1</i> chez le blé.....	281
2. Amplification de <i>MET1</i> chez <i>T. monococtum</i>.....	281
3. Crible des populations TILLING	283
3.1.1. Identification des codons stop potentiels	283
3.2. Design des amorces utilisées pour le crible des populations TILLING	283
3.3. Crible des populations.....	285
3.4. Validation des mutants.....	287
4. Etablissement et validation des lignées RNAi	287
4.1. Clonage	287
4.2. Transformation Agrobacterium	289
4.3. Transformation du blé et génotypage des événements.....	291
4.4. Suivi des générations des plantes issues de la 1ère vague RNAi.....	291
4.5. Suivi des générations des plantes issues de la 2ème vague RNAi.....	291
4.6. Recherche des sRNA par Northern Blot	293
5. Etude des profils de méthylation	295
5.1. Extraction de l'ADNg.....	295
5.2. Crible enzymatique des méthylations.....	295
5.3. Conversion au bisulfite.....	297
5.4. Design des amorces pour l'amplification sur ADN converti au bisulfite pour un séquençage MiSeq (Illumina).....	299
5.4.1. Séquences utilisées pour le design.....	299
5.4.2. Etablissement des profils CpG.....	299
5.4.3. Règles pour la définition des amorces	299
5.5. Amplification PCR sur ADN converti au bisulfite.....	301
Références	303
Annexes	319

Liste des abréviations

5mC : 5-méthylecytosine
 ADN : Acide DésoxyriboNucléique
 AGO : ARGONAUTE
 ARN : Acide RiboNucléique
 BAH : Bromo-Adjacent Homology
 Bal : *Brigadier x Alcedo*
 BLAST : Basic Local Alignment Search Tool
 cM : centiMorgan
 CMT2 : CHROMOMETHYLASE 2
 CMT3 : CHROMOMETHYLASE 3
 CO : Crossing Over
 Col : Columbia
 DAPI : 4',6'-diamidino-2-phénylindole
 DCL : DICER
 DDM1 : DECREASE DNA METHYLATION 1
 DME : DEMETER
 DNMT1 : DNA (cytosine-5)-METHYLTRANSFERASE 1
 DNMT1-RFD : DNMT1 Replication Foci Domain
 DNMT3 : DNA (cytosine-5)-METHYLTRANSFERASE 3
 DRM2 : DOMAIN REARRANGED METHYLTRANSFERASE 2
 EMS : Ethyl Methane Sulfonate
 EpiRILs : Epi Recombinant Inbred Line
 EST : Expressed Sequenced Tag
 ET/TE : Eléments Transposable/Transposable Elements
 EVD : EVADE
 FIS2 : FERTILIZATION INDEPENDENT SEED 2
 FT : FLOWERING LOCUS T
 HDA6 : HISTONE DEACETYLASE 6
 ICR : Imprinting Control Region
 IDN2 : INVOLVED IN DE NOVO 2
 IWGSC : International Wheat Genome Sequencing Consortium
 kb : kilobase
 Ler : Landsberg
 LINEs : Long Interspersed Nuclear Elements
 m6A : N⁶-methyladenine
 Mb : Megabase
 MEA : MEDEA
 MET1 : METHYLTRANSFERASE 1
 miRNA : microRNA
 MITEs : Miniature Inverted Transposable Elements
 Mt : Millions de tonnes
 NORs : Nucleolar Organizer Regions
 NRPE1 : NUCLEAR RNA POLYMERASE E1
 pb : paire de bases
 PCR : Polymerase Chain Reaction
 piRNA : piwiRNA

Ppd-1 : PHOTOPERIOD 1
 RT-qPCR : Reverse Transcription quantitative Polymerase Chain Reaction
 rasiRNA : repeat-associated siRNA
 RdDM : RNA-directed DNA METHYLATION
 RDR2 : RNA-DEPENDENT RNA POLYMERASE
 ReR : *Renan x Recital*
 Rht-1 : REDUCED HEIGHT 1
 RISC : RNA Induced Silencing Complex
 RNAi : RiboNucleic Acid interference
 SIM : Seed Inoculation Method
 siRNA/ARN : small interferent RNA/ARN
 SPTL5 : SUPPRESSOR OF TYINSERTION 5-LIKE
 sRNA : small RNA
 SUVH4 : SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE 4
 tasiRNA : trans acting RNA
 TILLING : Targeting Induced Local IN Genome
 TSS : Transcription Start Site
 UHRF1 : Ubiquitin-like containing PHD and RING Finger domain
 UTR : UnTranslated Region
 VIM : VARIANT IN METHYLATION
 Vrn-1 : VERNALIZATION 1
 Vrn-2 : VERNALIZATION 2
 Vrn-3 : VERNALIZATION 3
 WGD : Whole Genome Duplication
 WS : Wassilewskija
 WT : Wild Type

Liste des illustrations

Figure 1 : La production en blé tendre dans le monde et en France.....	16
Figure 2 : Demande mondiale en céréales.....	18
Figure 3 : Schéma simplifié de l’histoire évolutive du blé.....	20
Figure 4 : Variabilité des tailles de génome des plantes à fleurs.....	22
Figure 5 : Schématisation de l’organisation de la chromatine dans un noyau d’ <i>Arabidopsis</i>	28
Figure 6 : Noyau de souris et d’ <i>Arabidopsis</i> à l’interphase.....	30
Figure 7 : Caractéristiques des quatre états chromatinien principaux chez <i>Arabidopsis</i>	36
Figure 8 : Mise en place et maintien des 5mC chez la souris.....	44
Figure 9 : Modèle de l’empreinte parentale (imprinting) chez <i>Arabidopsis</i>	52
Figure 10 : Schématisation du RNA-directed DNA methylation, RdDM.....	58
Figure 11 : Maintient de la méthylation de l’ADN au cours de la réplication.....	60
Figure 12 : Représentation schématique du gène <i>MET1</i> chez le riz : <i>OsMET1a</i>	62
Tableau 1 : Liste non exhaustive de mutants touchant le gène <i>MET1</i> chez les plantes.....	64
Figure 13 : Etablissement des EpiRIL <i>met1</i> et <i>ddm1</i> chez <i>Arabidopsis thaliana</i>	76
Figure 14 : Principales phases du projet de thèse.....	78
Figure 15: Arbre phylogénétique retraçant l’histoire évolutive des espèces ayant un ancêtre commun avec le blé.....	82
Figure 16: Représentation schématique des ESTs des gènes <i>MET1</i> et <i>DDM1</i> identifiées chez le blé tendre....	82
Figure 17: Découverte de nouvelles séquences génomiques chez le blé par Sequence Capture.....	84
Figure 18 : Résultats de la première vague de Sequence Capture pour le gène <i>MET1</i> (expérience W05).....	86
Figure 19: Résultats combinés des deux vagues de Sequence Capture (SeqCap) pour le gène <i>MET1</i> (expérience W05 + W06).....	86
Figure 20: Analyse phylogénétique des contigs Sequence Capture.....	88
Figure 21 : Localisation des trois amplicons VRN1 utilisés pour la mise au point de la conversion au bisulfite.....	90
Figure 22: Schématisation de notre approche expérimentale.....	92
Figure 23: Histoire évolutive du génome du riz et du blé.....	96
Figure 24 : Analyse de la synténie entre le riz, <i>Brachypodium</i> et le blé dans la région des copies <i>TaMET-1</i> des chromosomes 5 et 7.....	216
Figure 25: Datation des deux duplications ayant conduit à la formation de trois paralogues de <i>MET1</i> chez les <i>Triticeae</i>	216
Tableau 2 : Bilan de la caractérisation des gènes <i>TaMET1</i> chez le blé tendre.....	220
Figure 26 : Stratégie de production d’une population de TILLING.....	222
Tableau 3 : Effectif des populations de TILLING chez le blé tendre hexaploïde disponibles à Biogemma ...	224
Tableau 4 : Nombre de codons CAA, CAG, TGG et CGA pour les quatre séquences utilisées lors du premier crible des populations TILLING.....	226
Figure 27 : Détection de polymorphisme par la méthode KASPAR.....	226
Figure 28: Principe théorique de la stratégie RNAi par formation d’ARN double brins en tige-boucle.....	228
Figure 29 : Principe de transformation du blé par la Seed Inoculation Method (SIM).....	228
Figure 30 : Représentation schématique du vecteur utilisé pour la première vague de RNAi contre <i>MET1</i> ...	230
Tableau 5 : Récapitulatif des générations de la première vague de RNAi contre <i>MET1</i>	230
Figure 31 : Résultats de recherche de siRNA par Northern Blot dans les lignées T1 exprimant un transgène RNAi dirigé contre <i>MET1</i>	232
Tableau 6 : Récapitulatif des semis des plantes T2 issus de la lignée T0 B053038.....	232

Figure 32 : Alignement de la séquence incluse dans le transgène RNAi « vague 2 » et les transcrits <i>TaMET1</i> des groupes 2 et 5.	234
Tableau 7 : Récapitulatif des générations de la deuxième vague de RNAi contre <i>TaMET1</i>	234
Figure 33 : Représentation schématique des lignées aneuploïdes utilisées.....	236
Tableau 8 : Récapitulatif du matériel végétal criblé (TILLING, lignées de délétion et mutants d’irradiation) ou produit (RNAi) au cours de mon projet de thèse.....	238
Figure 34 : Modèle d’interaction des gènes impliqués dans la floraison du blé.....	240
Figure 35 : Effet de plusieurs allèles du gène <i>Rht-1</i> (nanisme) sur la taille des tiges chez le blé	240
Figure 36 : Récapitulatif des régions promotrices <i>TaMET1</i> choisies pour l’étude des profils de méthylation	242
Figure 37 : Principe du crible d’identification de régions méthylées par digestion avec des enzymes sensibles à la méthylation.....	244
Figure 38 : Résultats du crible d’identification de régions méthylées par digestion par des enzymes sensibles à la méthylation pour le gène <i>Ppd-1</i>	244
Figure 39A : Résumé des résultats de crible de méthylation, des profils CpG et des zones choisies pour l’étude des profils de méthylation	246
Figure 39B : Résumé des résultats de crible de méthylation, des profils CpG et des zones choisies pour l’étude des profils de méthylation	248
Figure 41 : Résultats de PCR1 pour les gènes cibles	250
Figure 42: Résultats des PCR2 pour les gènes cibles.....	250
Figure 43 : Amplification PCR1 et PCR2 sur <i>Chinese Spring</i> pour la production des amplicons pour le séquençage MiSeq.....	252
Tableau 9 : Amorces utilisées pour l’amplification de <i>MET1</i> sur <i>T.monococcum</i>	280
Tableau 10 : Effectifs des plantes issues de la première vague de RNAi	292
Tableau 11 : Effectifs des plantes issues de la deuxième vague de RNAi	292
Tableau 12 : Oligonucléotides utilisés comme échelle externe pour les Northern Blot.....	294
Tableau 13 : Sondes utilisées pour l’hybridation des membranes.....	294
Tableau 14 : Conditions de digestion des enzymes sensibles à la méthylation.....	296
Figure 41 : Principe d’amplification par PCR imbriquée pour le séquençage MiSeq.....	300
Annexes.....	319
Mise au point de la méthode dite de « conversion au bisulfite » sur l’ADN génomique de blé tendre	321
Profil d’expression des gènes <i>MET1</i> chez <i>Arabidopsis</i> et le riz.....	339
Liste des amorces utilisées pour le premier crible des populations de TILLING	341
Représentation schématique du vecteur utilisé pour la deuxième vague de RNAi contre MET1	343
Amorces utilisées pour le crible de la méthylation des gènes cibles avec les enzymes sensibles à la méthylation	345
Amorces utilisées pour la production des amplicons pour le séquençage MiSeq	347

Synthèse bibliographique

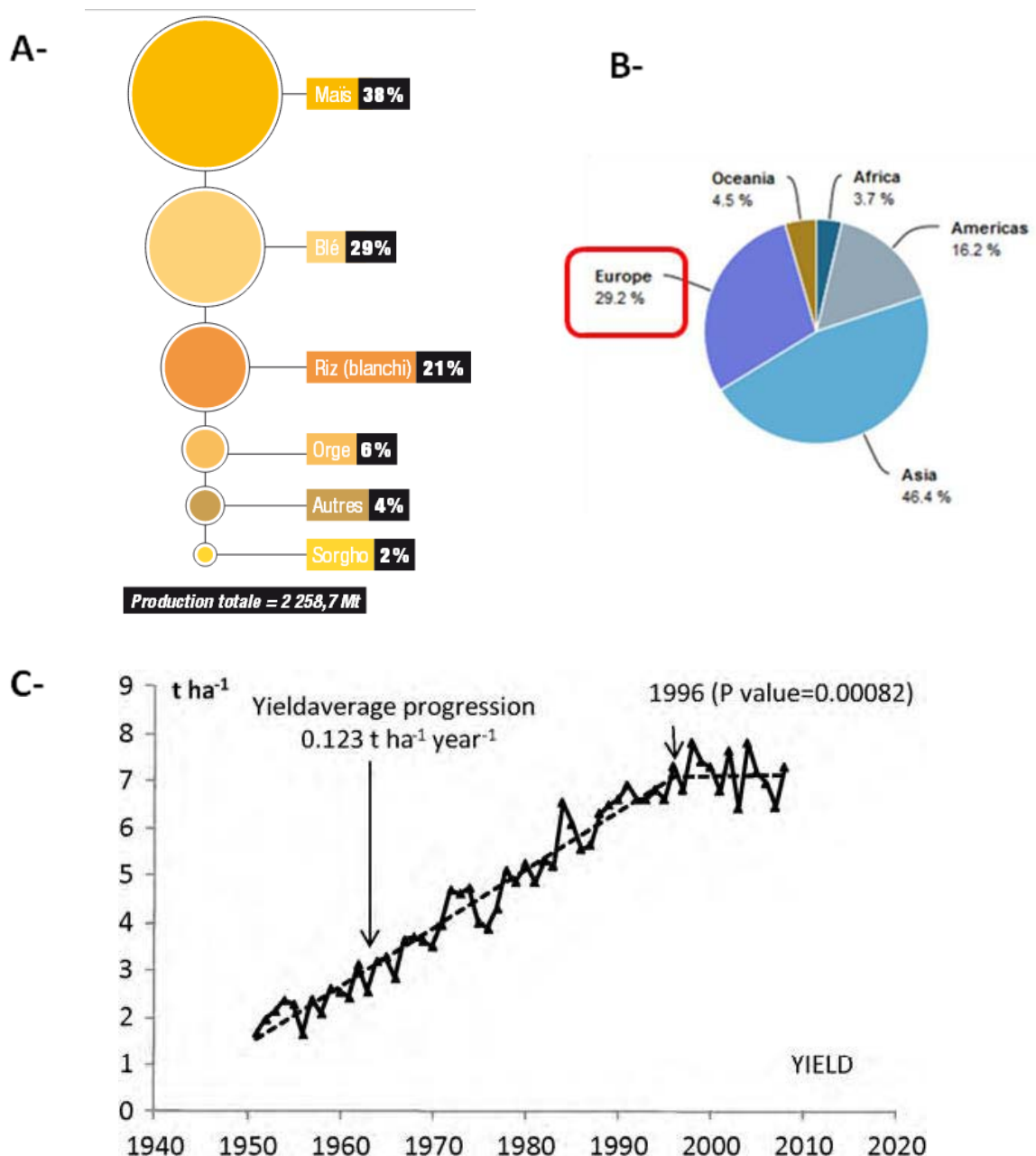


Figure 1 : La production en blé tendre dans le monde et en France.

A- Source : Passion céréale (www.passioncereales.fr)

La part des différentes productions céréalières est donnée en pourcentage de la production totale (exprimée en million de tonnes Mt) pour la saison 2012-2013. La production de blé représente 29% de la production totale en céréale.

B- Source : FAO (<http://faostat3.fao.org/>)

Le diagramme donne la part de chaque continent dans la production mondiale de blé tendre pour l'année 2012.

C- Source : (Brisson et al., 2010)

La courbe montre l'évolution des rendements de production de blé tendre en France. Les rendements (en tonne par hectare, t.h⁻¹) sont donnés entre les années 1950 et 2010. Un modèle statistique a été utilisé pour définir l'apparition d'un point d'inflexion en 1996. Depuis cette date, il y a une stagnation de la progression des rendements.

Chapitre 1 : le blé

Le blé est l'une des trois céréales majeures cultivées sur Terre avec le maïs et le riz. Le blé fait largement partie de notre vie de tous les jours. Une large fraction de la surface agricole en France est allouée à la production de blé, notamment en Auvergne. Le blé entre dans la composition des aliments consommés au quotidien à travers le pain ou les pâtes par exemple, produit respectivement à partir du blé tendre et du blé dur. L'importance de cette espèce dans l'alimentation humaine se retrouve aussi dans l'alimentation animale, en France mais également dans de nombreuses régions du monde.

1. Données économiques

En 2012-2013, la production mondiale de blé représentait 657 Millions de tonnes (Mt). Cela représente 30% de la production mondiale de céréale, soit la seconde après le maïs.

Le blé est un atout important pour la compétitivité de notre pays. En effet, la France est le premier pays producteur et exportateur de blé en Europe. Avec 35Mt produites en 2011, il s'agit de la première céréale cultivée en France et 54% de la production est exportée vers l'UE et le reste du monde. Les principales utilisations sont la meunerie et l'industrie alimentaire animale (<http://www.passioncereales.fr/>).

La France investit aussi beaucoup dans la recherche publique et privée visant à améliorer les variétés cultivées de blé. Mon projet de thèse s'inscrit dans un programme de recherche en partenariat entre deux de ces acteurs que sont le CNRS (Centre National de la Recherche Scientifique), un acteur public de la recherche française et l'entreprise privée Biogemma spécialisée dans la recherche sur les céréales pour le compte de plusieurs semenciers (Limagrain, RAGT et Euralis) et deux instituts techniques (Unigrain et Sofiproteol).

Grâce aux progrès techniques, à l'amélioration des méthodes agricoles et à la sélection génétique des variétés, les rendements de blé ont été multipliés par plus de trois depuis 1950, avec toutefois une stagnation des rendements observée depuis 1996 (Fig 1) (Brisson et *al.*, 2010). Cette stagnation n'est pas spécifique à la France, mais concerne l'ensemble des productions mondiales.

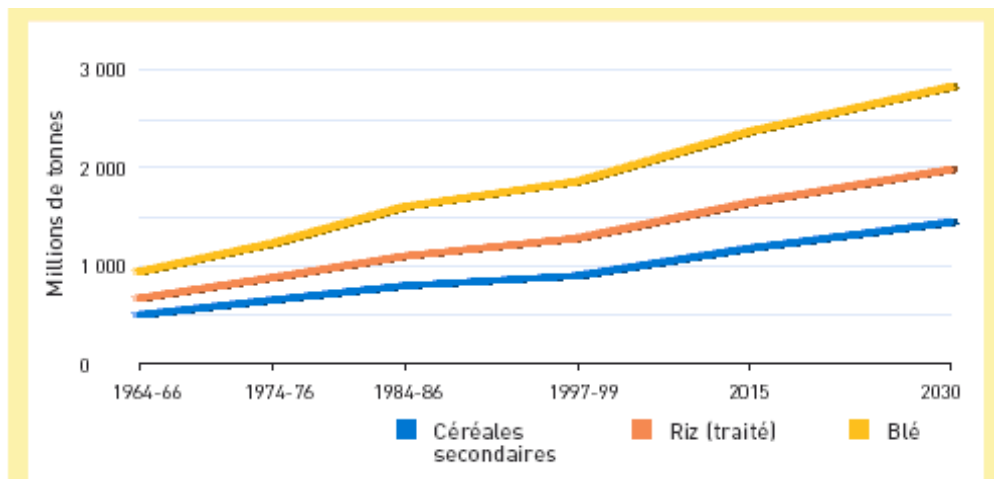


Figure 2 : Demande mondiale en céréales

Source : FAO (<http://www.fao.org/>)

La demande en céréale est donnée en millions de tonnes à partir de l'année 1965 et prédite jusqu'à l'année 2030 pour le blé, le riz et d'autres céréales.

L'une des constatations alarmantes de ces dernières années est l'observation que la population mondiale ne cesse d'augmenter et toutes les tendances prédisent une progression de la démographie d'ici à 2050 (<http://www.fao.org/docrep/016/ap106e/ap106e.pdf>). Ainsi, la prédiction de la demande en céréale dont celle du blé va donc également augmenter (Fig 2) (<http://www.fao.org/docrep/004/y3557f/y3557f08.htm>).

L'augmentation de la demande n'est pas le seul challenge pour la culture du blé. Les récents bouleversements climatiques tels que l'élévation des températures qui s'accompagne d'un raccourcissement des périodes de froid nécessaires à la floraison des cultures, de la modification des précipitations qui nécessitent une augmentation de l'irrigation des cultures, de la réduction des surfaces agricoles par l'urbanisation et l'augmentation du niveau des mers sont autant de contraintes qui doivent être prises en compte afin de créer de nouvelles variétés mieux adaptées à ces nouvelles conditions environnementales.

Enfin, parallèlement, ces nouvelles variétés doivent aujourd'hui être produites à rendement égal dans un contexte environnemental qui limite l'emploi des produits phytosanitaires (moins de pesticides ou d'herbicides, moins d'azote...etc). En effet la tendance actuelle prône un retour vers une agriculture raisonnée plus soucieuse de la protection de l'environnement tout en maintenant la productivité des cultures.

En conclusion, la demande en blé sera de plus en plus grande au vue des prévisions de croissance démographique alors que les surfaces cultivées ne seront pas étendues voire diminuées. De plus, les nouvelles variétés devront s'adapter aux changements climatiques et répondre aux nouvelles contraintes liées à l'agriculture moderne. La recherche sur le blé s'avère d'ores et déjà indispensable aux futurs gains de productivité nécessaires pour parvenir à nourrir l'ensemble des populations.

2. Histoire évolutive du blé tendre

Améliorer le blé, c'est en premier lieu retracer ses origines afin de mieux comprendre l'organisation de son génome.

Le blé tendre ou *Triticum aestivum* appartient au genre *Triticum*, et à la tribu des *Triticeae* tout comme l'orge. L'origine géographique du blé, ainsi que de l'orge, des pois et des lentilles

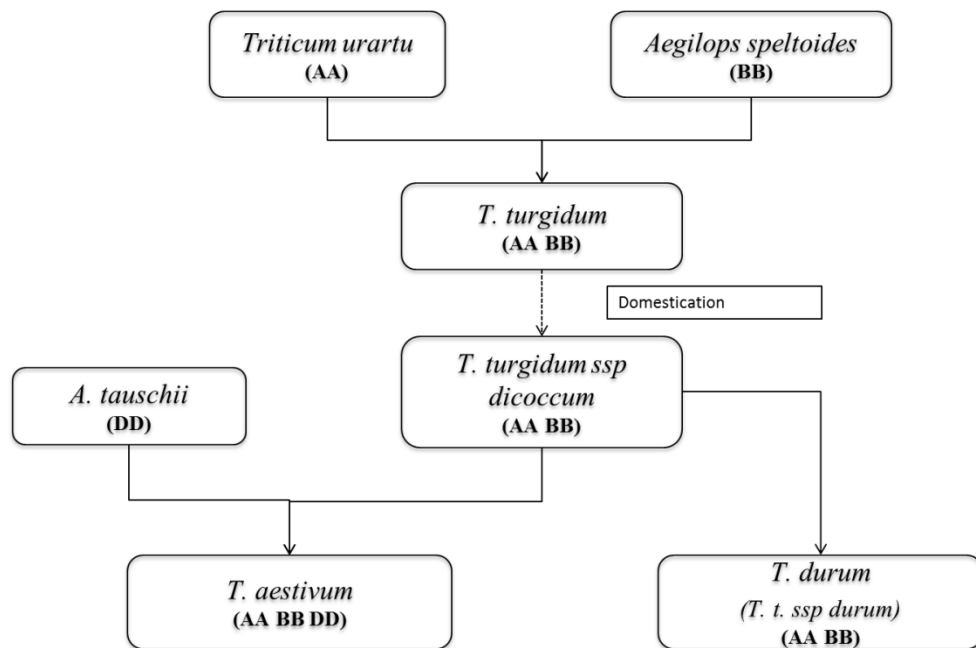


Figure 3 : Schéma simplifié de l'histoire évolutive du blé

Les génomes diploïdes de *T. urartu* (génome A) et *A. speltoides* (génome B) ont donné le tétraploïde *T. turgidum*. Ce dernier a été domestiqué en une sous espèce *T. turgidum ssp dicoccum*, qui s'est associée à *A. tauschii* (génome D) pour former le génome hexaploïde actuel du blé tendre, *T. aestivum*. *T. t. ssp dicoccum* est également à l'origine du blé dur actuel *T. durum*.

se situe au niveau du croissant fertile et plus précisément dans une région qui se trouve aujourd'hui à la frontière de la Turquie et de la Syrie (Lev-Yadun et *al.*, 2000). Le blé tendre possède un génome complexe de très grande taille. Il est hexaploïde, c'est à dire constitué de trois génomes diploïdes qui se sont assemblés par « polyploïdisation ». Toutefois, l'organisation actuelle du génome du blé tendre montre que d'autres événements sont survenus au cours du temps tels que la sélection naturelle, mécanisme qui permet l'évolution des espèces et l'adaptation à leurs milieux, et la domestication qui se traduit par l'influence de l'activité humaine sur certains caractères et comportements héréditaires. Chacun de ces génomes diploïdes est composé de sept paires de chromosomes ($2n = 14$). Les trois génomes ancestraux sont qualifiés de « génomes homéologues ».

Il y a 0,25 à 1,3 Ma, *Triticum urartu* (génome AA) s'est associé avec une espèce proche d'*Aegilops speltoides* aujourd'hui disparue pour donner une sous-espèce de *T. turgidum* tétraploïde, de génome AA BB (Fig 3). Cette espèce a ensuite été domestiquée et a évolué pour donner *T. turgidum ssp dicoccum* l'ancêtre du blé dur (*T. durum*). Cet ancêtre tétraploïde s'est associé il y a 8 000 ans avec *A. tauschii* (génome DD) donnant le blé hexaploïde actuel de génome AA BB DD ($2n = 42$) (Charmet, 2011).

3. Le génome du blé

3.1. Comparaison des trois génomes homéologues du blé tendre

Le blé tendre est constitué de trois génomes diploïdes très proches, qui dérivent d'un même ancêtre commun. De ce fait, la plupart des gènes se retrouvent en théorie présents en trois exemplaires au sein du génome hexaploïde. Afin de mieux comprendre la structure hexaploïde, des études de « génomique comparative » (ou synténie) ont été réalisées sur de grandes régions des chromosomes 7A, 7B et 7D (Berkman et *al.*, 2013). La comparaison a été réalisée avec d'autres céréales diploïdes telles que le riz, *Brachypodium* et le sorgho qui partage un ancêtre commun diploïde avec le blé. Ainsi, cette étude permet de mieux définir la conservation de l'ordre des gènes sur un segment donné entre deux espèces ayant un ancêtre commun.

Au niveau des régions non synténiques, c'est-à-dire celles ayant subi des réarrangements chromosomiques, chacun des trois génomes compte 50% de gènes qui lui sont spécifiques. Par contre dans les régions synténiques, seulement 4.9%, 6.6% et 11.2% de ces

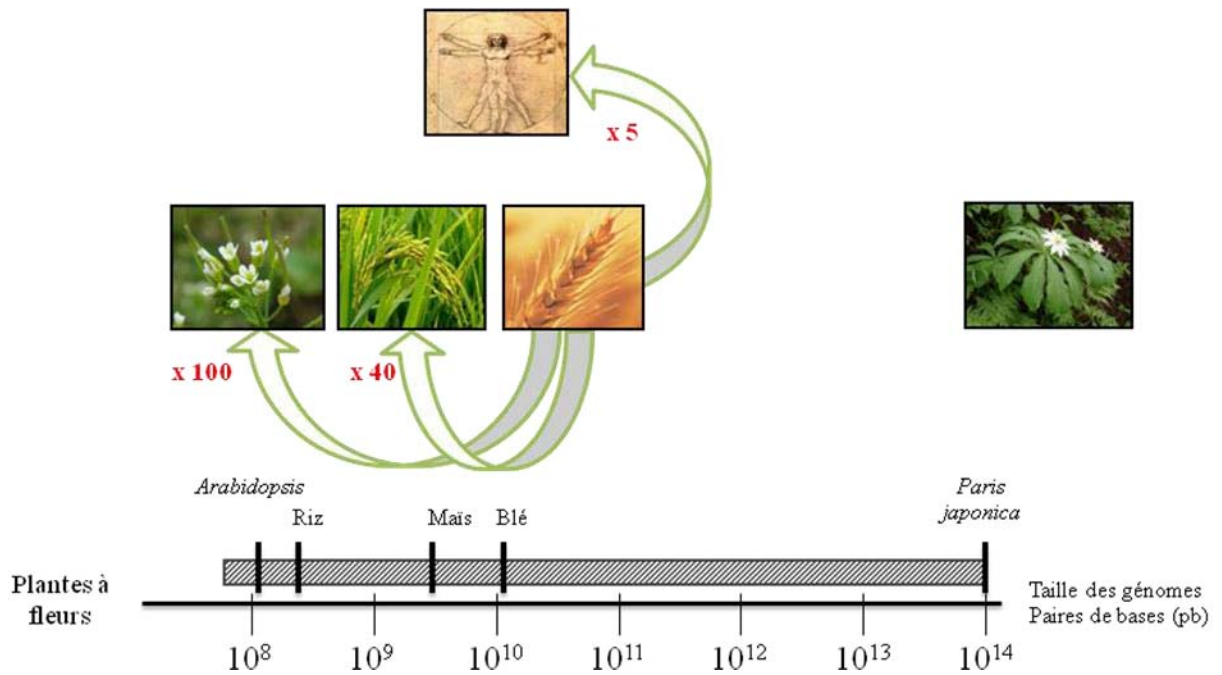


Figure 4 : Variabilité des tailles de génome des plantes à fleurs.

L'axe horizontale représente la taille des génomes (exprimés en paires de bases pb). L'ordre de grandeur des génomes d'*Arabidopsis*, du riz, du maïs et du blé sont indiqués sur cet axe, ainsi que celui de *Paris Japonica*, plante à fleur possédant le plus grand génome. Les rapports de taille du génome du blé avec ceux d'*Arabidopsis*, du riz et à titre indicatif celui de l'Homme, sont indiqués en rouge.

gènes sont spécifiques respectivement au génome A, B et D. Cette constatation implique qu'une perte de gènes a eu lieu pendant ou après la polyploïdisation et que cette perte a été plus importante pour les génomes A et B. En plus des différences dans le nombre de gènes perdus entre les génomes homéologues, ainsi que la nature des gènes impliqués diffèrent. Pour les génomes A et B, ce sont principalement des réseaux de gènes qui codent pour des protéines ribosomiques, des facteurs de transcription et des ARN polymérasés. Bien que les gènes du chromosome 7D soient moins largement impliqués dans des réseaux, ce chromosome est enrichi en gènes dont la fonction est liée à la prise en charge des ions métalliques et au développement reproductif.

En conclusion, malgré le fait que le génome du blé hexaploïde soit composé de trois génomes homéologues tous issus d'un même ancêtre commun et donc initialement avec la même organisation génomique, la période suivant la polyploïdisation a conduit à une réorganisation importante. Un phénomène de perte de gènes est observé, associé à une accumulation de Single Nucleotide Polymorphism (SNP) pour les gènes conservés, de 3 à 4,5 fois plus nombreux dans les génomes A et B comparés au génome D (Berkman *et al.*, 2013).

3.2. Organisation générale du génome du blé

Le génome du blé tendre fait partie des grands génomes dans le règne végétal avec près de 17 000Mb (Fig 4). La totalité de la séquence du génome du blé n'est pas encore disponible mais c'est le cas pour le chromosome 3B. En effet, le chromosome 3B, qui est le plus grand des chromosomes du blé (~1 000Mb), peut-être purifié du reste des chromosomes par cytométrie de flux, permettant son séquençage (Safár *et al.*, 2004). Par extrapolation des données obtenues pour le chromosome 3B, le nombre de gènes pour la totalité du génome B du blé tendre serait compris entre 36 000 et 50 000 gènes (Choulet *et al.*, 2010). Cette valeur est légèrement supérieure au nombre de gènes présent dans les génomes du riz, du sorgho ou du maïs (International Rice Genome Sequencing, 2005 ; Paterson *et al.*, 2009 ; Schnable *et al.*, 2009). Ces gènes ont une taille très variable avec une moyenne de 3300 ± 2900 pb. En moyenne, la densité de gènes est de l'ordre de 1 gène tous les 104kb, avec une densité plus importante au niveau des extrémités du chromosome, avec 1 gène pour 86kb, que dans la partie proximale (proche du centromère) avec 1 gène pour 184kb. La plupart de ces gènes sont regroupés en petits clusters comptant moins de 4 gènes (3.2 ± 1.6 gènes). Seules les régions autour du centromère semblent montrer une absence de

gène, mais à l'inverse présentent une forte proportion d'éléments transposables.

Les éléments transposables représentent un peu plus de 80% du génome du blé tendre. Les éléments les plus présents sont les rétrotransposons à LTR, en particulier les éléments de type Gypsy et Copia qui représentent près de 50% du génome à eux seuls. Les transposons de type CACTA sont également fortement représentés. Les éléments transposables se retrouvent à la fois dans les régions intergéniques mais également aux niveaux des gènes. Ainsi, 33% des gènes comportent dans leurs introns des éléments transposables qui dans ce cas, sont des éléments de type Miniature Inverted Transposable Elements (MITEs) et Long Interspersed Nuclear Elements (LINEs).

Ainsi, la grande taille du génome du blé s'explique par le fait que le blé est un organisme hexaploïde ayant subi de nombreux réarrangements chromosomiques et une invasion de son génome par les éléments transposables (Choulet et *al.*, 2010).

3.3. Le séquençage du génome du blé

Comme nous l'avons vu précédemment, les demandes en blé sont de plus en plus importantes, pourtant les rendements stagnent. Les apports des méthodes agricoles arrivent à leurs limites, les changements climatiques et les attentes sociétales sont de plus en plus contraignantes. Il y a donc une demande importante vis-à-vis de la création de nouvelles variétés pour améliorer les performances du blé et l'adapter à son nouvel environnement. Cette stratégie repose actuellement soit sur des démarches sans *a priori* comme la sélection génomique soit sur des démarches plus ciblées comme la sélection assistée par marqueurs, notamment en suivant des allèles de gènes ayant un intérêt agronomique ou encore *via* la transgénèse en introduisant un nouveau gène ou en surexprimant/sousexprimant un gène de l'espèce considérée (Varshney et *al.*, 2005 ; Anderson, 2007). Concernant les démarches ciblées, le séquençage du génome semble donc une stratégie très prometteuse afin d'identifier un grand nombre de nouveaux marqueurs et de nouveaux gènes.

Du fait de sa grande taille, de la présence de trois génomes homéologues et de sa forte proportion en éléments transposables, le génome du blé est l'une des dernières espèces d'intérêt agronomique majeur dont la séquence du génome n'est pas disponible. En 2013, les séquences de deux des génomes diploïdes ont été publiées: celui de *T. urartu* (génome A) et de *A. tauschii* (génome D). Le génome diploïde ancêtre du génome B n'est toujours pas

séquencé. Il en est de même pour le génome hexaploïde qui devrait toutefois être publié au cours de l'année 2014. En effet, afin de disposer de la première séquence du génome du blé tendre, un grand effort de séquençage a été fait à travers le consortium de séquençage appelé International Wheat Genome Sequencing Consortium (IWGSC) impliquant des acteurs publics et privés de différentes parties du monde (<http://www.wheatgenome.org/>).

Il s'agit d'un séquençage de chacun des bras chromosomiques purifiés par cytométrie de flux. A ce jour, tous les bras chromosomiques sont séquencés. Bien qu'il s'agisse d'une quantité d'information importante, ces séquences couvrent essentiellement les zones géniques, donc une fraction encore limitée du génome du blé. Toutefois, le chromosome 3B est pour sa part entièrement séquencé et assemblé. Les chromosomes 6B et 7B sont également très complets. Les chromosomes 1A, 4A, 7A, 1B et 7D ont une séquence de référence uniquement dans les parties distales des chromosomes (<http://www.wheatgenome.org/>).

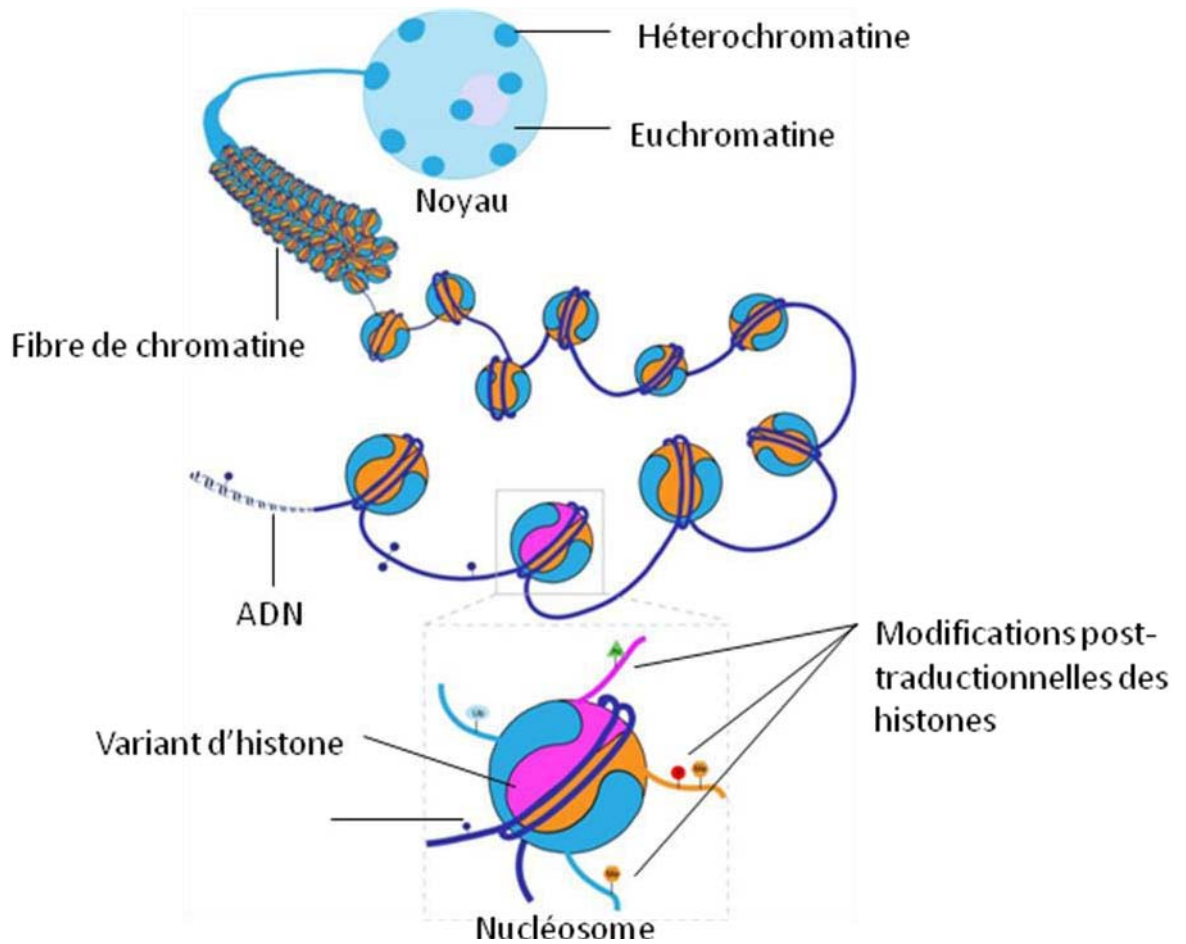


Figure 5 : Schématisation de l'organisation de la chromatine dans un noyau d'Arabidopsis.

L'unité de base de la chromatine est le nucléosome formé de l'enroulement de la molécule d'ADN autour d'un octamère d'histone. Ce nucléosome peut subir des modifications post-traductionnelles. Les nucléosomes vont subir différents niveaux de compaction pour former la fibre de chromatine. Cette fibre de chromatine sera présente sous deux formes plus ou moins compactées au sein du noyau : l'hétérochromatine et l'euchromatine.

Chapitre 2 : épigénétique & méthylation de l'ADN

Jusqu'ici l'amélioration des variétés s'est essentiellement basée sur la génétique, par exemple en utilisant la sélection assistée par marqueurs, la transgénèse ou la sélection génomique. Pourtant la génétique, c'est-à-dire la variabilité de la séquence d'ADN, n'explique pas la totalité de la performance d'une variété notamment son aptitude à répondre aux changements environnementaux (sécheresse, froid, carence en nutriments...). La recherche a découvert ces dernières années d'autres modifications du génome dites « épigénétiques ». Afin de bien comprendre ce type de modification, il convient de rappeler que, chez les eucaryotes, l'ADN n'est pas nu mais organisé en chromatine. Grâce à des protéines et des ARN, la chromatine permet d'une part la compaction de l'ADN afin qu'il puisse être contenu dans le noyau d'une cellule, et d'autre part la régulation transcriptionnelle du génome. La chromatine porte des « marques épigénétiques » qui vont définir l'état d'activité transcriptionnelle : les marques épigénétiques permissives permettent de définir les séquences transcriptionnellement actives alors que les marques répressives se retrouvent au niveau des séquences réprimées du génome. L'épigénétique se définit comme les modifications héréditaires de l'expression du génome indépendantes de la séquence d'ADN.

1. Les différents niveaux de compaction de la chromatine

Du fait de la grande taille des génomes, la chromatine se retrouve sous forme compactée dans le noyau (Fig 5). Le niveau de compaction peut varier en fonction du cycle cellulaire. 147 paires de bases (pb) de la double hélice d'ADN (diamètre estimé à $\sim 2\text{nm}$) s'enroulent autour d'un complexe protéique pour former le nucléosome qui est considéré comme l'unité de base de la chromatine. Ce complexe protéique est constitué d'un octamère d'histones : 2 molécules d'histones H2A, H2B, H3 et H4 (Luger et *al.*, 1997). Chacune des histones, exceptée l'histone H4, existe sous plusieurs formes (variants d'histones) codées par des gènes différents. La chromatine forme une structure dite en collier de perles de 10 nm de diamètre. Les nucléosomes peuvent s'enrouler en hélice pour former une fibre d'organisation supérieure de 30 nm de diamètre.

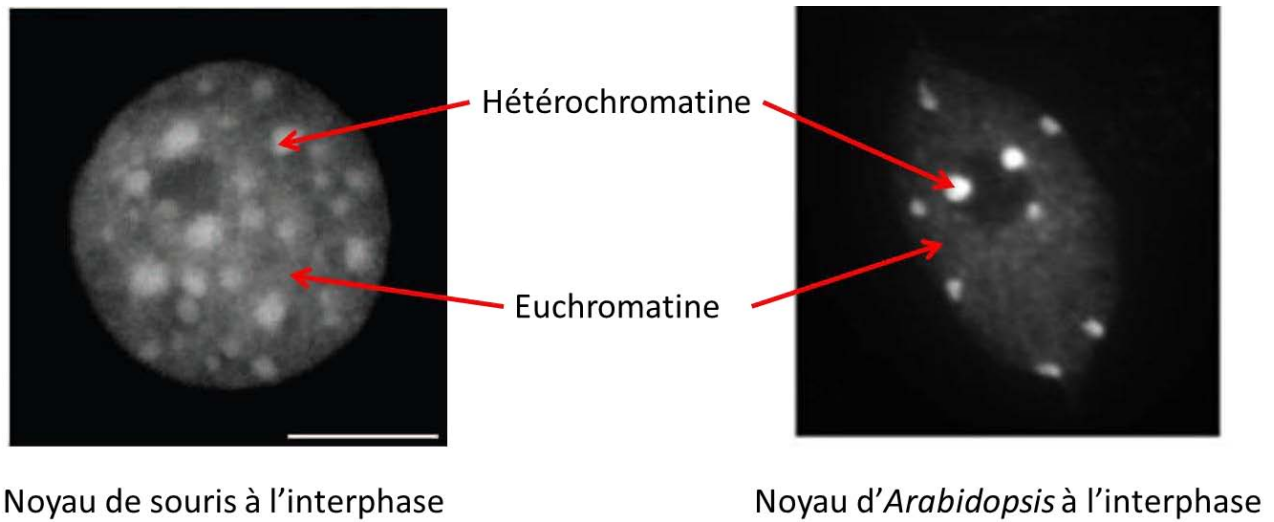


Figure 6 : Noyau de souris et d'*Arabidopsis* à l'interphase

Source : (Probst et *al.*, 2009) & (Benoit et *al.*, 2013)

Les noyaux sont marqués avec du 4',6'-diamidino-2-phénylindole (DAPI), un intercalant de l'ADN. Les régions de fortes intensités correspondent à l'hétérochromatine, état condensé de la chromatine. Ces régions bien visibles en microscopie sont également appelées des « chromocentres ». Les régions d'intensités plus diffuses sont constituées d'euchromatine plus décondensée.

La formation de cette fibre nécessite l'intervention d'autres protéines dont l'histone H1. C'est à ce niveau que s'exercerait le rôle régulateur des histones faisant passer la fibre de chromatine d'un état permissif pour la transcription (fibre de 10nm) à un état répressif (fibre de 30nm) et inversement. Toutefois, comme nous le verrons plus tard des modifications post-traductionnelles des histones telles que la méthylation ou l'acétylation sont également impliquées dans l'établissement des états permissifs ou répressifs de la chromatine de 10nm sans pour autant la modifier en fibre de 30 nm. La fibre chromatinienne peut subir un niveau de compaction supérieur en une fibre de 300 nm qui forme des boucles de chromatine attachées à la matrice nucléaire. Ces boucles pourraient également participer à la définition de domaines de chromatine d'ordre supérieur. Enfin, une dernière condensation permet de former une fibre de 700nm retrouvée uniquement dans les chromosomes métaphasiques. C'est le niveau de compaction le plus important connu à ce jour (Fussner et *al.*, 2011 ; Bian and Belmont, 2012).

1.1. Euchromatine et hétérochromatine

Il n'est pas toujours facile de définir les deux principaux états de la chromatine, car selon les auteurs plusieurs critères peuvent être employés : la compaction, la position cytologique, les marques épigénétiques, l'enrichissement en gènes ou en séquences répétées ou encore l'activité transcriptionnelle. J'ai choisi d'illustrer les deux états de la chromatine selon un critère cytologique qui est celui qui a été historiquement décrit en premier par Heitz (Heitz, 1928). Les noyaux peuvent en effet facilement être colorés avec du 4',6'-diamidino-2-phénylindole (DAPI), un agent intercalant de l'ADN capable de fluorescer sous lumière UV, permettant de visualiser l'organisation de la chromatine au sein du noyau. Une telle méthode employée sur des noyaux à l'interphase de souris et d'*Arabidopsis thaliana* révèle des zones d'intensités de fluorescence différentes (Fig 6). Les régions de forte intensité correspondent à la chromatine sous sa forme compacte, répressive, appelée hétérochromatine. Les régions d'intensité plus diffuse correspondent à la chromatine relâchée, appelée euchromatine (Probst et *al.*, 2009 ; Benoit et *al.*, 2013 ; Vanrobays et *al.*, 2013).

L'euchromatine et l'hétérochromatine ne sont pas distribuées de façon aléatoire le long des chromosomes. Chez *Arabidopsis*, l'euchromatine est localisée sur les bras chromosomiques alors que l'hétérochromatine se retrouve concentrée dans plusieurs régions. Les régions centromériques s'organisent en hétérochromatine qui dans ce cas est constituée d'éléments répétés

tels que la séquence satellite de 180pb, et les éléments répétés appelés séquence 106B. Les régions péri-centromériques sont également hétérochromatiques et constituées pour une grande part d'éléments transposables et de gènes d'ARN ribosomiques (ARNr) 5S. L'hétérochromatine se retrouve également sur les bras chromosomiques intercalée entre deux régions euchromatiques (knobs), au niveau des Nucleolar Organizer Regions (NORs) qui portent les gènes d'ARNr 45S et des télomères (Fransz et *al.*, 2000).

En résumé, l'euchromatine est un état relâché de la chromatine riche en gènes alors que l'hétérochromatine est plus compactée et riche en séquences répétées.

1.2. Organisation de la chromatine en fonction de la taille des génomes

Selon la taille des génomes, l'organisation de la chromatine diffère au sein des noyaux. Selon certains auteurs, il convient de distinguer 3 types de génomes (van Driel and Fransz, 2004). Les petits génomes qui ont une taille inférieure à 500Mb, comme *Arabidopsis*, la *Drosophile* ou le riz ; les génomes de taille moyenne, comprise entre 500 et 3 000Mb, comme c'est le cas pour la souris, l'humain, la tomate; le maïs et le sorgho ; alors que les grands génomes ont une taille supérieure à 3 000Mb comme par exemple le blé ou l'orge.

Chez *Arabidopsis*, la fraction riche en éléments transposables et pauvre en gènes représente seulement 10% du génome et est localisée principalement au niveau des régions centromériques et péri-centromériques. A l'interphase, ces régions se regroupent au niveau des « chromocentres » (Fig 4) d'où émanent des boucles d'euchromatine contenant les régions riches en gènes. Cette organisation est appelée le « modèle en rosette » (Fransz et *al.*, 2002). Chez le riz, l'Homme et le sorgho les régions hétérochromatiques se retrouvent au niveau centromérique et péri-centromérique, mais également dispersées le long des chromosomes. Tout comme pour *Arabidopsis*, les zones hétérochromatiques pour ces trois organismes sont visibles en coloration DAPI sous forme de chromocentres, mais ceux-ci sont plus nombreux et le contraste entre zones condensées et décondensées est moins marqué (Cheng et *al.*, 2001 ; She et *al.*, 2007) (Fig 4). Chez l'Homme et la souris, les régions pauvres en gènes et les gènes réprimés sont associés avec la périphérie nucléaire et colocalisent avec l'hétérochromatine (van Driel and Fransz, 2004 ; Geyer et *al.*, 2011). Enfin, pour les grands génomes comme le blé ou l'orge, les régions riches en éléments répétés et les clusters de gènes sont présents tout le long du chromosome. A l'interphase, ces deux espèces ont une organisation dite « Rabl », c'est-à-dire que les bras chromosomiques

sont alignés et les centromères et les télomères sont à la périphérie nucléaire mais à des pôles opposés (She et *al.*, 2007 ; Santos and Shaw, 2004 ; Jasencakova et *al.*, 2001). Il est à noter que la configuration en Rabl ne dépend pas que de la taille des génomes. En effet, chez la levure *Saccharomyces cerevisiae* ou dans certains tissus du riz, les chromosomes adoptent aussi cette configuration (Santos and Shaw, 2004 ; Sáez-Vásquez and Gadal, 2010). Il n'y a pas encore d'explication claire démontrant pourquoi un organisme ou un type cellulaire adopte ce type de conformation.

En résumé, petits et moyens génomes concentrent leur hétérochromatine en position centromérique et pericentromérique, avec parfois un début d'extension le long des bras chromosomiques. D'un point de vue cytologique, l'hétérochromatine est visualisable au niveau des chromocentres qui pourraient être importants pour définir les boucles d'euchromatine. Pour les plus grands génomes, les éléments répétés semblent avoir « colonisés » tout le chromosome, il sera donc plus difficile de clairement définir des régions eu- et hétérochromatiques. Ceci pose des questions quant à la régulation transcriptionnelle des gènes dans les grands génomes, car ils se retrouvent dans un contexte globalement hétérochromatique.

2. Chromatine et différents niveaux de régulation épigénétique

Différents types de marques épigénétiques vont définir si une séquence est dans un état autorisant ou non son expression transcriptionnelle. Plusieurs niveaux de régulation épigénétique ont été décrits à ce jour telles que la méthylation de l'ADN, les modifications post-traductionnelles des histones, les variants d'histones, les petits ARN (small RNA, sRNA), le remodelage de la chromatine et la position de la chromatine dans le noyau ou architecture nucléaire.

La méthylation de l'ADN, qui sera plus longuement détaillée dans les paragraphes suivants, correspond à une modification des cytosines sur la molécule d'ADN. Schématiquement, les cytosines méthylées (5-methyl Cytosine, 5mC) sont le plus souvent associées à des régions compactées de la chromatine et donc silencieuses d'un point de vue transcriptionnel.

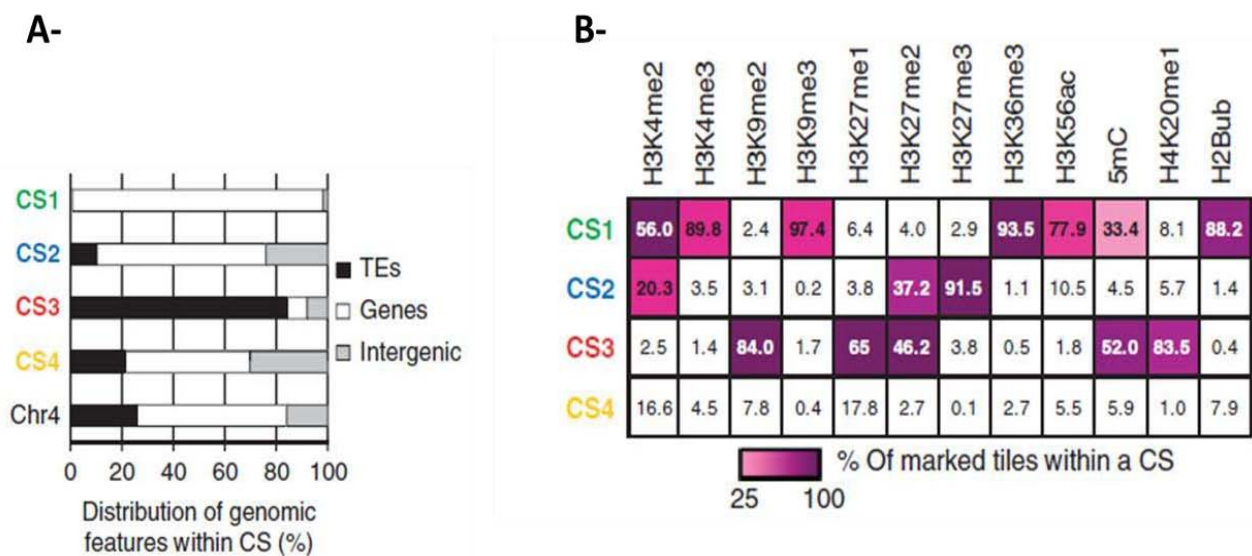


Figure 7 : Caractéristiques des quatre états chromatiniens principaux chez *Arabidopsis*

Source : (Roudier et *al.*, 2011)

Les quatre états chromatiniens (Chromatine State) sont appelés CS1, CS2, CS3 et CS4.

A- La figure donne la fraction d'éléments transposables (Transposable Element, TE), de gènes et de séquences intergéniques pour les quatre états chromatiniens et pour le chromosome entier (chromosome 4).

B- Le tableau donne la distribution des douze marques épigénétiques étudiées pour les quatre états chromatiniens principaux.

Une des marques épigénétiques la plus importante concerne les histones, les sous unités constitutives des nucléosomes. Les queues N-terminales des histones ainsi que leur partie globulaire peuvent subir des modifications post-traductionnelles, essentiellement au niveau des arginines et des lysines, telles que la méthylation (me), l'acétylation (ac), l'ubiquitination (ub), mais beaucoup d'autres modifications sont découvertes (Tan et *al.*, 2011). Ces modifications, propres à chaque histone voire à chaque variant d'histone sont collectivement regroupées sous le nom de « code histone », et sont associées à un état permissif ou répressif de la chromatine. Les méthylations (mono, di et tri methylation respectivement me1, me2 et me3) sur la lysine (K) 4 de l'histone H3 (H3K4me1/me2/me3) sont, associées avec les gènes et donc principalement concentrées dans l'euchromatine chez *Arabidopsis* et l'Homme. A l'inverse, la marque H3K9me2 est une marque répressive associée avec l'hétérochromatine chez *Arabidopsis* et l'Homme ; associée à la marque H3K9me3 uniquement chez l'Homme (Feng and Jacobsen, 2011).

L'étude de la répartition des modifications d'histones et de la méthylation de l'ADN ont permis de définir quatre états de la chromatine, correspondant chacune entre 19 et 31% du génome d'*Arabidopsis* (Roudier et *al.*, 2011). Ainsi, 12 marques épigénétiques, 11 marques sur les histones et la méthylation de l'ADN sur les cytosines (5mC) ont été choisies car connues pour être associées avec des activités transcriptionnelles particulières ou des localisations sub-nucléaires chez *Arabidopsis* (Fig 7). L'un de ces états chromatinien est composé de séquences géniques, intergéniques et d'éléments transposables (ETs) dans des proportions proches. Il ne présente pas d'enrichissement particulier dans aucune des 12 marques épigénétiques (CS4). Le second état chromatinien (CS1) est associé à l'euchromatine et aux gènes fortement exprimés: ces domaines sont enrichis en marques permissives H3K4me2, H3K36me3 et H2Bub. Un troisième état (CS2) est composé essentiellement de gènes qui sont faiblement exprimés et de quelques ETs et sont enrichis en marques H3K27me2 et H3K27me3. Enfin, le dernier état (CS3) ne concerne que l'hétérochromatine et est associé aux marques répressives H3K9me2, H3K27me1, H4K20me1 et 5mC (Roudier et *al.*, 2011).

Une découverte importante de ces 30 dernières années chez le nématode et les plantes (Baulcombe, 2004 ; Fire et *al.*, 1998) a été la mise en évidence des petits ARN (sRNA) qui sont aussi une composante importante pour établir un état répressif des séquences. Schématiquement, chez les plantes il existerait deux grands types de sRNA. D'une part, les microRNA (miRNA), trans acting RNA (tasiRNA) et small interfering RNA (siRNA) vont entraîner la

dégradation ou le blocage de la traduction d'un ARNm. D'autre part, les repeat-associated siRNA (rasiRNA) encore appelé heterochromatic RNA (hcRNA) provoquent la méthylation du gène cible (Eamens *et al.*, 2008). Ces derniers sont utilisés pour guider la méthylation de l'ADN grâce au mécanisme appelé RNA-directed DNA Methylation (RdDM) qui implique des ARN polymérase spécifiques des plantes appelées ARN polymérases IV et V (PolIV et PolV). Les rasiRNA sont également appelés plus couramment siRNA, et c'est ainsi qu'ils seront nommés quand la voie RdDM sera décrite dans le paragraphe 3.3.2.1..Notons qu'une troisième classe de sRNA appelé piwiRNA (piRNA) a été décrite chez les animaux mais n'a pas été retrouvée chez les plantes bien que leur mode d'action rappelle le mécanisme mis en jeu lors du RdDM (Law and Jacobsen, 2010). Quelque soit le type de sRNA, certaines étapes conduisant à leur formation sont communes. En effet, ces mécanismes forment tous à un stade des ARN double brins (ARNdb) qui seront pris en charge par un complexe protéique comprenant entre autre des protéines de type DICER (DCL). Ce complexe va cliver les longs ARNs double brins en sRNAs, qui seront pris en charge par les protéines de la famille ARGONAUTE (AGO) (Eamens *et al.*, 2008). Selon le type de voie privilégiée, les sRNAs vont soit cibler les ARNm, il s'agit alors de posttranscriptional gene silencing (PTGS), soit induire une méthylation de l'ADN, il s'agit alors de transcriptional gene silencing (TGS) (Fagard and Vaucheret, 2000).

Les régulations épigénétiques font intervenir un grand nombre d'acteurs qui influencent l'expression du génome en favorisant la transcription des gènes, ou au contraire en permettant la répression de la fraction répétée et transposable du génome. Ces deux aspects sont particulièrement intéressants chez le blé tendre. En effet, la plus grande proportion de son génome correspond à des ET, qui ont la faculté de bouger et d'induire des mutations. Des clusters de gènes dont l'expression est nécessaire à son développement sont intercalés dans ces larges fractions d'hétérochromatine. Le maintien de la stabilité du génome est assuré par l'ensemble de ces marques épigénétiques. Nous avons choisi de nous concentrer sur un de ces mécanismes : la méthylation de l'ADN.

3. La méthylation de l'ADN

3.1. La méthylation chez les organismes vivants

3.1.1. Les méthylations de l'ADN.

La méthylation de l'ADN se retrouve au niveau de trois bases : N⁶-methyladenine (m6A), 5-methylcytosine (m5C) et N⁴-methylcytosine. Toutes ces méthylations sont mises en place par des enzymes de type DNA methyltransferases après la réplication de l'ADN. Chez les eucaryotes, seules les m6A et les m5C ont été répertoriées, cependant les m5C sont de loin les plus répandues au sein des génomes. Les m6A ont été mises en évidence de façon directe par chromatographie chez plusieurs organismes appartenant aux groupes des protistes, chez certains champignons, le blé et le moustique, ainsi que par digestion enzymatique chez d'autres organismes de ces mêmes groupes (Ratel et al., 2006).

Pour la suite, lorsqu'il sera fait référence à la méthylation de l'ADN il s'agira des méthylations de type m5C. L'étude de la méthylation (méthylome) chez l'ensemble des eucaryotes montre que certaines espèces ont perdu l'aptitude à méthyler les m5C ou le font de façon très inefficace. En effet, il n'y a pratiquement pas de méthylation chez le nématode *Caenorhabditis elegans* (Simpson et al., 1986) alors qu'elle existe chez un autre nématode, *Trichinella spiralis* (Gao et al., 2012). Le génome de *Drosophila melanogaster* ne comporte qu'un niveau très faible de méthylation (Lyko et al., 2000) et bien qu'initialement la levure ne semblait pas posséder de m5C dans son génome, des traces ont récemment été décrites (Tang et al., 2012). Certaines espèces présentent également un taux de méthylation très faible comme l'abeille *Apis mellifera* (Feng et al., 2010).

Même à très faible niveau, la méthylation de l'ADN serait donc universelle et présente des procaryotes aux eucaryotes. Certains auteurs décrivent cette modification épigénétique comme la 5^{ème} base de l'ADN (Lister and Ecker, 2009).

3.1.2. La méthylation de l'ADN chez les animaux et les végétaux.

Les trois contextes de méthylation : CpG, CpHpG et CpHpH

La méthylation des cytosines en contexte CpG (Cytosine-phosphate-Guanine) est conservée chez les plantes et les animaux. Dans le génome humain, 80% des cytosines sont méthylées dans ce contexte, ce qui représente 1% du génome (Ehrlich et al., 1982). Chez la souris, le niveau de méthylation est proche de celui du génome humain puisque près de 75%

des cytosines en contexte CpG sont méthylées (Feng et *al.*, 2010). Chez les plantes, les cytosines sont également préférentiellement méthylées en contexte CpG mais le niveau de méthylation apparaît plus variable. Par exemple, chez *Arabidopsis* seules 24% des cytosines sont méthylées dans ce contexte (Cokus et *al.*, 2008) alors que les CpG sont méthylés à plus de 40% chez le peuplier et à près de 60% chez le riz (Feng et *al.*, 2010).

Les cytosines en contexte CpHpG et CpHpH peuvent aussi être méthylées. Ce type de méthylation est bien décrit chez les plantes (Gruenbaum et *al.*, 1981 ; Meyer et *al.*, 1994) : en effet 6,7% et 1,7% des cytosines sont méthylées respectivement en contexte CpHpG et CpHpH chez *Arabidopsis* (Cokus et *al.*, 2008). La méthylation en contexte non CpG est également retrouvée chez les animaux mais de façon moins systématique. Seuls certains types de tissus sont concernés, en particulier les cellules souches embryonnaires où 25% des cytosines en contexte non CpG peuvent être méthylées. Ces méthylations sont perdues au cours de la différenciation et pourraient donc jouer un rôle essentiellement dans le développement précoce de l'embryon (Lister et *al.*, 2009).

La répartition des cytosines au sein du génome est également variable entre les génomes animaux et végétaux. Chez la souris, le pourcentage de CpG méthylés varie peu (entre 70% et 80%) entre les régions géniques et non géniques. A l'inverse, chez les plantes, le pourcentage de méthylation des régions répétées est variable selon le contexte de méthylation. Il est de 2 à 10 fois plus important que celui des régions géniques (Feng et *al.*, 2010).

Chez l'Homme, dans un contexte de méthylation plutôt uniforme du génome, des zones de 200 à 1 000pb présentant un enrichissement en CpG et associées à une absence de méthylation sont retrouvées : il s'agit des « îlots CpG ». Près de 50% de ces îlots sont retrouvés au niveau des promoteurs et des sites d'initiation de la transcription (Transcription Start Site, TSS). Les 50% restant sont appelés des « îlots orphelins », présents en position intra ou intergénique. Toutefois, des analyses à l'échelle du génome ont montré que 40% d'entre eux correspondaient à des sites d'initiation de la transcription non annotés. Les îlots présents au niveau des promoteurs ainsi que les îlots orphelins, qu'ils soient en position intra- ou intergénique, sont enrichis en H3K4me3 qui est une marque épigénétique associée aux promoteurs actifs (Deaton and Bird, 2011 ; Illingworth et *al.*, 2010).

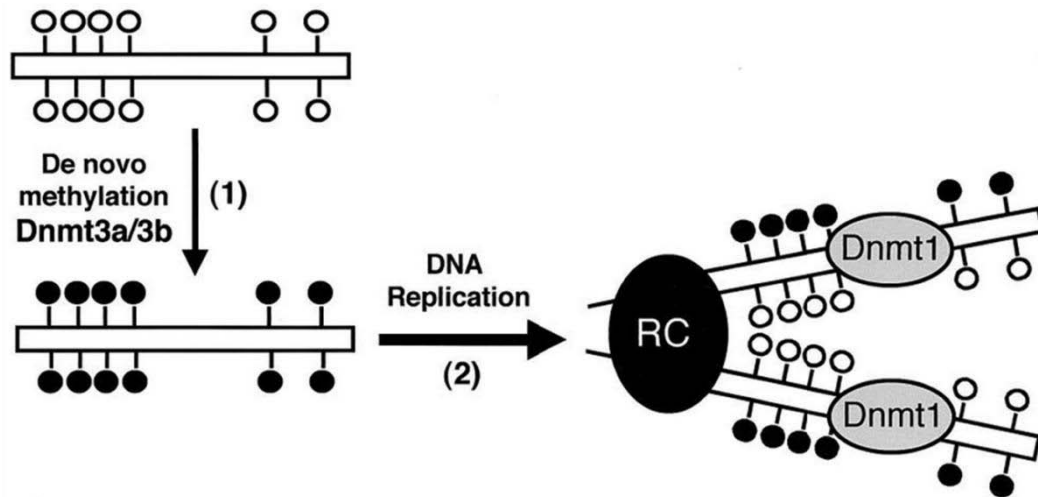


Figure 8 : Mise en place et maintien des 5mC chez la souris

Source : (Chen et *al.*, 2003)

Les cytosines non méthylées, indiquées par des ronds blancs, sont méthylées (ronds noirs) par Dnmt3a/3b (étape 1) puis maintenues pendant la réplication par Dnmt1 (étape 2).

Methylation de novo et de maintenance

Chez les plantes et les animaux, il existe une voie de mise en place de la méthylation et une voie de maintien de la méthylation. Chez les plantes, le maintien de la méthylation est assuré en contexte CpG par MET1 (homologue de DNMT1 chez les animaux), par CMT3 en contexte CpHpG et par CMT2 en contexte CpHpH (voir paragraphe 3.3). Il existe cependant des différences dans les voies de mise en place des méthylations entre les animaux et les plantes.

Chez les animaux, la mise en place des méthylations est assurée par DNMT3A et DNMT3B (Fig 8). Des souris mutées pour *dnmt3a* mènent leur développement à terme mais meurent rapidement ensuite ; celles mutées pour *dnmt3b* ou doublement mutées *dnmt3a dnmt3b* ne sont pas viables. DNMT3A et B n'ont pas d'influence sur le maintien de la méthylation, qui est assuré par DNMT1 (Fig 6) (Okano et al., 1999). Ces deux enzymes mettent en place une grande partie de la méthylation très tôt au cours de l'embryogénèse ; s'en suit une vague de déméthylation pour effacer « les profils de méthylation » de la génération précédente. Les profils de méthylation sont ensuite rétablis par DNMT3A qui est recruté par la marque d'histone H3K4 (Law and Jacobsen, 2010). Chez les plantes, la mise en place de la méthylation se fait grâce aux sRNA et à des polymérases spécifiques. En effet, c'est la voie du RdDM qui n'existe pas chez les animaux qui va déterminer les zones à méthyler en utilisant des sRNA comme guides. Le RdDM sera décrit plus précisément au paragraphe 3.3.2.1.

3.1.3. Déméthylation active de l'ADN

Chez les plantes, il existe quatre glycosylases qui sont responsables de la déméthylation active de l'ADN. La première est DEMETER (DME), une glycosylase fortement exprimée dans les tissus reproducteurs femelles avant fécondation (Choi et al., 2002). DME agit en excisant les 5mC au niveau de l'allèle maternel du gène *MEDEA* (*MEA*) et assure ainsi son activation transcriptionnelle nécessaire au développement normal des graines (Gehring et al., 2006). La protéine codée par le gène *REPRESSOR OF TRANSCRIPTIONAL GENE SILENCING*, ROS1, possède également une activité glycosylase. Elle a été identifiée chez un mutant *Arabidopsis* montrant la répression d'un transgène luciférase actif ainsi que du gène endogène ayant un promoteur homologue. Cette répression est due à une hyperméthylation du promoteur et elle peut être levée par l'utilisation d'un agent déméthylant, la 5-azacytidine (Gong et al., 2002). Enfin, deux protéines homologues de ROS1 et DME, DEMETER-LIKE 2 et 3

(DML2 et DML3) sont nécessaires pour le maintien d'un statut hypométhylé de certaines cibles, comme le rétrotransposons à LTR *AtGPI* ou les extrémités 5' et 3' de nombreux gènes (Ortega-Galisteo et al., 2008 ; Penterman et al., 2007). Ces deux gènes ont un profil d'expression moins étroit que DME puisqu'ils sont exprimés dans les feuilles, les fleurs, les tiges et les racines (Ortega-Galisteo et al., 2008).

Chez les mammifères, il existe deux voies de déméthylation active. Dans une première voie, l'ACTIVATION-INDUCED DEAMINASE (AID) déamine en premier les 5mC qui seront ensuite excisées par la THYMINE DNA GLYCOSYLASE (TDG) (Cortellino et al., 2011). La seconde voie fait d'abord intervenir la protéine TET1 qui va oxyder les 5mC en 5hydroxy-mC. Celles-ci pourront redevenir des cytosines par un mécanisme non connu ou être prises en charge par AID (Guo et al., 2011).

3.2. Rôles de la méthylation

3.2.1. Méthylation et augmentation de la mobilité des ETs

Un des rôles important de la méthylation de l'ADN est le maintien dans un état réprimé des ETs.

Chez *Arabidopsis*, des mouvements d'un rétrotransposon appelé EVADE (EVD) ont été observés lors d'une étude de la descendance issue de la lignée homozygote mutante *met1-3*. Cette mutation provoque une diminution importante de la méthylation globale du génome. La diminution de la méthylation en contexte CpG est également observée au niveau de la région promotrice d'EVD et provoque sa dérégulation transcriptionnelle puis sa transposition. La mobilité d'EVD induit des insertions dans des gènes et se traduit par des phénotypes mutants, comme par exemple l'absence d'organes floraux masculins et/ou l'absence de pétales (Mirouze et al., 2009).

Les modifications importantes du méthylome engendrées par le mutant *met1-3* ont été mises à profit afin de créer des populations de lignées présentant des altérations épigénétiques ou « EpiRILs » (Epi Recombinant Inbred Line). A partir d'un croisement entre le mutant homozygote *met1-3* et une plante sauvage Col 0, la mutation *met1-3* a été contre-sélectionnée dès la génération F2 (voir Objectifs, Fig 10). A ce stade, chaque plante F2 est autofécondée jusqu'à la huitième génération (F8). Les lignées EpiRILs montrent un taux plus faible de méthylation du génome comparé au sauvage mais globalement les profils de méthylation « perdus » dans la plante homozygote *met1-3* sont restaurés au cours des 8 générations (Reinders et al., 2009).

Les régions du génome où l'on retrouve le plus d'ET, les régions centromériques et péri-centromériques, sont d'ailleurs les premières cibles de la reméthylation limitant ainsi leur mobilité. Toutefois, des transpositions de transposons de types CACTA ont été répertoriées dans ces lignées, transpositions qui s'accumulent au cours des générations. EVD n'est donc pas le seul élément transposable à être mobilisé dans un mutant *met1* (Reinders et al., 2009).

Chez les mammifères, la méthylation de l'ADN est importante dans le fonctionnement normal des cellules *via* la répression des ETs. Des souris hétérozygotes pour un allèle hypomorphe du gène *DNMT1* sont viables et présentent une hypométhylation du génome. Ces souris montrent des signes d'instabilité du génome ainsi que des événements de transpositions d'éléments rétroviraux endogènes. L'insertion d'un de ces éléments a été observée au niveau du locus *Notch1*. Cette insertion dérégule l'expression de *NOTCH1* et provoque l'apparition de tumeurs (Howard et al., 2007).

3.2.2. Méthylation et transcription

Chez l'homme, le méthylome est souvent modifié dans les cellules cancéreuses en comparaison aux cellules normales. L'hypométhylation générale du génome observée dans les cellules cancéreuses s'aggrave lors du passage de l'état de prolifération bénigne à l'état de cancer invasif. A l'inverse, les îlots CpG, méthylés dans les cellules normales, deviennent hypométhylés dans les cellules cancéreuses (Esteller, 2008). Cette hyper-méthylation peut concerner des gènes suppresseurs de tumeur et donc entraîner leur répression transcriptionnelle. C'est le cas du gène *p16* impliqué dans le contrôle du cycle cellulaire. L'hyperméthylation d'un îlot CpG au niveau de son promoteur est corrélée avec sa répression transcriptionnelle puisque le traitement des cellules cancéreuses avec un agent bloquant la méthylation, l'azacytidine, induit une reprise de sa transcription. La méthylation du promoteur de *p16* régule donc directement son expression (Herman et al., 1995).

Chez la plante, l'exemple le plus démonstratif est sans doute celui du gène *FWA* d'*Arabidopsis*. *FWA* est un gène impliqué dans la floraison et une forte expression de *FWA* induit un phénotype de floraison tardive. Pour fleurir, *FWA* doit être réprimé et cette répression est mise en place grâce à des mécanismes épigénétiques impliquant la méthylation de l'ADN (Kinoshita et al., 2004). Des éléments répétés présentant des homologies avec des Short Interspersed Nuclear Element (SINE) sont retrouvés au niveau du TSS, et ces petites séquences répétées en amont du TSS sont méthylées. Chez des plantes présentant une diminution

de la méthylation comme dans un mutant *met1* ou *ddm1*, ces éléments répétés sont déméthylés. Ceci provoque une levée de la répression de *FWA* et cause un retard de floraison (Soppe et al., 2000 ; Kinoshita et al., 2007 ; Lippman et al., 2004). D'autres gènes tels que *SUPERMAN* et *AGAMOUS* subissent quant à eux une hyperméthylation chez ces mêmes mutants *met1* et *ddm1*. Dans ce cas, ces gènes sont réprimés et entraînent des anomalies de développement floral (Jacobsen et al., 2000).

Une étude de la méthylation de l'ADN par immunoprécipitation des 5mC menées conjointement avec une étude du transcriptome chez *Arabidopsis* a permis d'établir les relations entre expression, transcription et méthylation. Les gènes les plus fortement et les plus faiblement exprimés font partie des gènes les moins méthylés. Globalement, la méthylation des gènes a un effet répresseur sur l'expression. Chez les mutants *met1-6*, il y a une surexpression de l'ensemble des gènes en comparaison avec une plante sauvage. Cette surexpression est plus importante pour les gènes qui sont normalement méthylés dans la plante sauvage : la déméthylation est donc en partie directement responsable de leur dérépression (Zilberman et al., 2007).

La distribution des 5mC au niveau des gènes n'est pas aléatoire. En effet, l'ensemble des gènes méthylés montrent une absence de méthylation dans les premiers 2kb (proche du TSS) et le dernier 1kb des gènes. Ce profil est l'inverse de celui retrouvé pour la fixation de l'ARN polymérase II (Pol II), qui est enrichie aux extrémités 5' et 3' et qui marque les transitions initiation-élongation et terminaison-polyadénylation. Ceci montre que la présence de méthylations en début et fin de gène interfère avec la transcription. La méthylation de l'ADN dans le corps des gènes va également interférer avec l'élongation de la transcription. Ainsi, les gènes les plus surexprimés dans le mutant *met1-6* sont ceux qui présentent la plus forte méthylation du corps du gène en contexte sauvage. L'effet répresseur de la méthylation est particulièrement efficace pour les gènes de petite taille (inférieur à 2,5kb) car ce sont eux qui sont les moins exprimés parmi les gènes méthylés (Zilberman et al., 2007). Les relations existant entre méthylation, expression et taille des gènes ont ainsi permis de déduire que les gènes les moins méthylés sont impliqués dans des fonctions ribosomiques, structurales ou sont des facteurs de transcription. Les gènes les plus méthylés quant à eux ont des rôles de liaison aux nucléotides, aux acides nucléiques et sont impliqués dans des activités de type kinase (Zilberman et al., 2007).

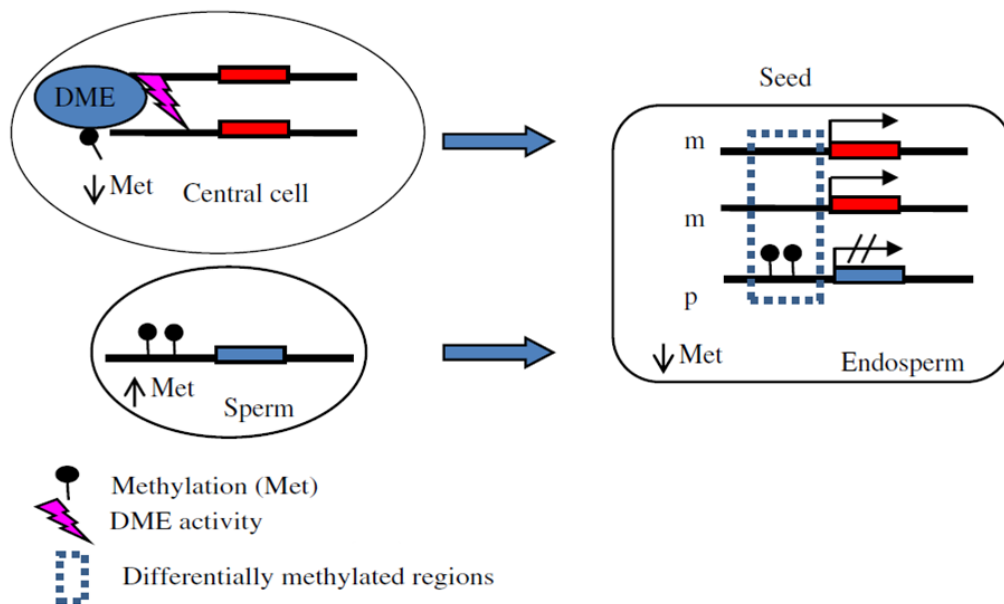


Figure 9 : Modèle de l’empreinte parentale (imprinting) chez *Arabidopsis*.

Adapté de (Zhang et al., 2013)

Le gène DEMETER (DME) est actif dans la cellule centrale (central cell) et provoque la déméthylation d’un gène cible (en rouge). A l’inverse, l’absence d’activité de DME dans la cellule spermatique (sperm) autorise le maintien de la méthylation dans le promoteur du gène cible (en bleu). Après fécondation, il y a une expression de l’allèle maternel (m) mais pas de l’allèle paternel (p). Ce modèle montre l’impact de la méthylation de l’ADN sur l’expression différentielle des allèles parentaux.

Un autre lien entre méthylation de l'ADN et expression des gènes est l'empreinte parentale, ou *imprinting* en anglais. Il s'agit d'un mécanisme permettant de distinguer les deux allèles parentaux *via* des marques épigénétiques. Les gènes à empreinte parentale s'expriment de façon mono-allélique et spécifique à l'une ou l'autre des origines parentales. La marque épigénétique de base impliquée dans ce mécanisme est la méthylation de l'ADN (Fig 9). Une centaine de gènes sont concernés par l'empreinte parentale chez les mammifères. La mise en place de la méthylation différentielle entre les deux allèles parentaux et donc leur expression différentielle est établie dans les premiers stades de développement et maintenue tout au long de la vie (Weaver et *al.*, 2009). Une perte de l'empreinte parentale peut avoir de graves conséquences comme c'est le cas dans le syndrome de Prader-Willi chez l'Homme, qui se traduit par des effets sur le développement néonatal et pendant l'enfance (Bervini and Herzog, 2013).

Chez les plantes, le nombre de gènes soumis à l'empreinte est mal défini. Selon les études réalisées sur *Arabidopsis* il y aurait entre 30 et 165 gènes à expression maternelle et de 20 à 43 gènes à expression paternelle (Hsieh et *al.*, 2011 ; Wolff et *al.*, 2011 ; Gehring et *al.*, 2011 ; Zhang et *al.*, 2013). Comparativement, une étude chez le riz identifie 180 gènes à expression maternelle et 85 gènes à expression paternelle (Luo et *al.*, 2011). Parmi les quelques gènes connus chez *Arabidopsis* pour être soumis à l'empreinte parentale, citons *MEA*, *FWA* et *FIS2* dont les allèles parentaux seront différentiellement exprimés grâce à l'action de la déméthylase DEMETER (Zhang et *al.*, 2013) (Fig 9).

3.2.3. Méthylation et recombinaison

La fréquence moyenne des crossing over (CO) chez *Arabidopsis* est similaire entre les bras chromosomiques (3,55cM/Mb) et les régions péri-centromériques (3,83cM/Mb) avec une forte augmentation pourtant à la limite entre ces deux régions (Yelina et *al.*, 2012). Dans les régions péri-centromériques, la fréquence des COs est positivement corrélée avec la densité en gènes et négativement avec la méthylation de l'ADN ; de telles corrélations n'existent pas dans les régions des bras chromosomiques (Yelina et *al.*, 2012). Des croisements effectués entre le mutant *met1-3* dans un fond génétique Columbia (Col *met1-3*), l'écotype Columbia (Col) et l'écotype Landsberg (Ler) (croisements Col x Ler ou Col *met1-3* x Ler) ont permis d'étudier le lien entre la perte de la méthylation et la fréquence de COs. Ler et Col sont polymorphes et il sera ainsi possible de distinguer les allèles Col de ceux de Ler (Yelina et *al.*, 2012). Dans les lignées issues du croisement Col *met1-3* x Ler, le pourcentage des COs comparée au croisement

Col x Ler (fixé par référence à 100%) est de 140% dans les régions euchromatiques et diminue à 80% dans les régions hétérochromatiques (Mirouze et *al.*, 2012). L'étude utilisant des croisements entre écotypes sauvages (Col x Ler) et entre mutant et sauvage (Col *met1-3* x Ler) confirme ces résultats et indique que bien qu'il y ait une diminution de la fréquence des COs dans les régions péri-centromériques, cette fréquence augmente dans les régions centromériques (Yelina et *al.*, 2012). Les deux analyses montrent qu'à l'échelle du génome la perte de la méthylation n'entraîne pas de modification dans le nombre de COs mais leur redistribution le long du chromosome (Yelina et *al.*, 2012 ; Mirouze et *al.*, 2012).

Il semble exister une répartition différente des COs dans le génome humain. Notons en premier lieu qu'il s'agit d'un génome de taille moyenne (~3Gb) alors que l'exemple développé ci-dessus chez *Arabidopsis* concerne un petit génome (~0,15Gb). Nous avons vu précédemment que la répartition des séquences répétées est différente entre ces deux types de génomes (voir paragraphe 1.2.). Ainsi, le génome humain peut être divisé en zones de forte recombinaison (3cM/Mb ou plus) et des zones de "deserts de recombinaison" (moins de 0,3cM/Mb) (Buard and de Massy, 2007). Les régions sujettes aux COs sont regroupées en clusters. Dans ce cas, il n'existe que des preuves indirectes d'une implication éventuelle des marques épigénétiques sur le taux de recombinaison. Les analyses montrent en effet que le taux de recombinaison augmente avec le contenu en G+C et la richesse en CpG. Il existe une corrélation négative entre la présence d'un hotspot de recombinaison et la présence d'un îlot CpG. D'une façon générale, les hotspots de recombinaison sont situés plutôt en dehors des gènes. Une observation pour l'instant inexplicée montre toutefois que les gènes soumis à l'empreinte parentale ont un taux de recombinaison plus élevé (Buard and de Massy, 2007).

3.3. Mise en place et maintien des profils de méthylation chez les plantes

3.3.1. Contexte de méthylation, proportion et répartition

Au sein des génomes des plantes, le profil de méthylation n'est pas uniforme. La méthylation des cytosines est plus importante au niveau des régions répétées du génome que dans les gènes, quelque soit le contexte de méthylation et l'espèce végétale étudiée.

Ceci a été bien décrit chez la plante modèle *Arabidopsis thaliana* (Lister et al., 2008). De manière générale, les séquences répétées sont plus méthylées que les gènes codant pour les protéines et leur méthylation est plus uniforme : à titre d'exemple, 91% des transposons sont méthylés alors que seuls 20% des gènes le sont (Zilberman et al., 2007). A l'échelle du génome, le pourcentage de méthylation en contexte CpG dans les gènes est de l'ordre de 25%, 50% et 30% respectivement pour *Arabidopsis*, le riz et le peuplier et augmentent à des taux de 75%, 85% et 90% si seules les séquences répétées sont considérées. Un profil similaire est observé pour les cytosines dans les contextes CpHpG et CpHpH de méthylation (Feng et al., 2010 ; Cokus et al., 2008).

La méthylation des gènes n'est pas aléatoire. En effet, chez *Arabidopsis*, les gènes proches des centromères sont plus souvent méthylés. Il existe une corrélation entre l'augmentation de la présence d'ETs et l'augmentation du nombre de gènes méthylés (Zilberman et al., 2007). Chez le riz, deux tiers des gènes associés à un ET sont méthylés alors que seuls 30% des gènes actifs et des pseudogènes le sont. Lorsqu'ils sont méthylés, les gènes peuvent être méthylés sur plus de 10% de leur longueur (Yan et al., 2010). La méthylation qui concerne dans ce cas presque exclusivement les contextes CpG, est relativement basse aux extrémités 3' et 5' et inversement, est associée à une augmentation graduelle pour former un pic de méthylation dans le corps du gène. Ce profil est indépendant de la taille du gène (Zilberman et al., 2007). Enfin, une légère diminution de la méthylation au niveau des introns est observée dans le corps du gène, cette diminution étant assez marquée chez *Arabidopsis* (Feng et al., 2010). Des observations relativement similaires ont été faites chez le riz et le peuplier.

Il existe donc une différence significative entre la méthylation des séquences répétées qui lorsqu'elles sont méthylées, le sont sur la totalité de leur séquence par rapport aux gènes. Dans le cas des séquences répétées, la méthylation peut s'étendre aux régions immédiatement adjacentes et pourrait donc influencer l'expression des gènes au voisinage. Les différences de profils des différents contextes entre les zones géniques et répétées peuvent s'expliquer par les mécanismes de mise en place des méthylations qui sont développés ci-dessous.

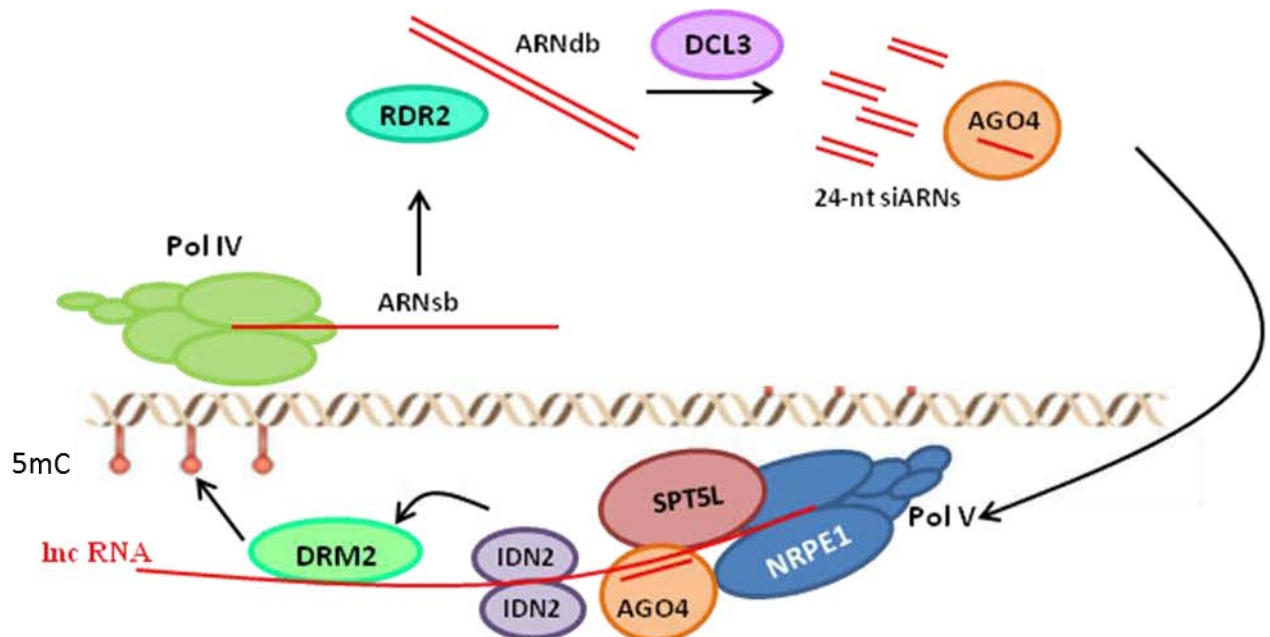


Figure 10 : Schématisation du RNA-directed DNA methylation, RdDM

Adapté de (Law and Jacobsen, 2010)

L'initiation du RdDM débute par la production d'un ARN simple brin (ARNsb) généré par une polymérase spécialisée des plantes (Pol IV). Un ARN double brin (ARNdb) est ensuite produit par la RNA-DEPENDENT RNA POLYMERASE (RDR2). DICER-LIKE 3 (DCL3) clive cet ARNdb en petits ARNs interférents (siARN) de 24 nucléotides, qui seront ensuite pris en charge par ARGONAUTE 4 (AGO4). AGO4 interagit avec la sous-unité NRPE1 d'une seconde polymérase spécifique des plantes (Pol V). Pol V est à l'origine de transcrits intergéniques non codants (INC, ou Long Intergenic Non Coding RNA: LincRNA) qui seront reconnus par les siARN de 24-nt. SUPPRESSOR OF TY INSERTION 5-LIKE (SPT5L) semble stabiliser la reconnaissance entre LincRNA et siRNA. Les hybrides LincRNA/siRNA recruterait INVOLVED IN DE NOVO 2 (IDN2) qui conjointement à SPT5L recruterait DOMAIN REARRANGED METHYLTRANSFERASE 2 (DRM2), principale méthyltransférase dans la voie RdDM. DRM2 induirait au final la méthylation de la séquence cible.

3.3.2. Mise en place et maintien des méthylations

3.3.2.1. La voie RNA-directed DNA Methylation (RdDM)

Chez la plante, la mise en place *de novo* de la méthylation se fait par la voie RdDM pour RNA-directed DNA Methylation, faisant intervenir des petits ARNs (Fig 10).

La première étape est la génération des petits ARNs (siRNAs). Tout d'abord, la RNA Polymerase IV (Pol IV), spécifique des plantes, va produire des ARNs simple brin (ARNsb). Puis la RNA-DEPENDENT RNA POLYMERASE (RDR2) va prendre en charge ces ARNs et les convertir en ARN double brin (ARNdb) qui seront ensuite découpés en siARN de 24 nucléotides (nt) par DICER 3 (DCL3). En parallèle, des transcrits issus des séquences intergéniques non codantes sont générés par l'action des polymérase Pol II et Pol V. L'accumulation de ces ARNs se fait indépendamment de l'action de DCL3 et RDR2. Ils auraient une action dans le recrutement des partenaires protéiques impliqués dans le RdDM (fonction de « matrice support »). Les siARNs sont pris en charge par le complexe AGO4, qui interagit avec la sous unité NUCLEAR RNA POLYMERASE E1 (NRPE1) de Pol V. Cette interaction est l'une des étapes clef nécessaire pour le RdDM (Law and Jacobsen, 2010).

Les siARNs et les transcrits produits par la Pol V sont également des acteurs indispensables. La liaison entre les deux voies se fait *via* SUPPRESSOR OF TY INSERTION 5-LIKE (SPT5L) qui interagit avec AGO4 et NRPE1. SPT5L pourrait être impliquée dans le recrutement de DRM2 (DOMAIN REARRANGED METHYLTRANSFERASE 2). Un autre facteur serait INVOLVED IN DE NOVO 2 (IDN2), dont l'un des substrats pourrait être les ARNdb formés entre les siARNs et les transcrits Pol V-dépendants. Tout comme SPT5L, IDN2 pourrait recruter DRM2 pour la mise en place de la méthylation (Zhang and Zhu, 2011 ; Law and Jacobsen, 2010). DRM2 est la principale méthyltransférase agissant dans la voie RdDM. En effet, l'altération d'un acide aminé indispensable à l'activité catalytique de la protéine DRM2 entraîne une absence de méthylation et ce quelque soit le contexte CpG, CpHpG ou CpHpH (Naumann et *al.*, 2011).

Le maintien des méthylations se fait au moment de la réplication de l'ADN et de ce fait, les mécanismes mis en jeu seront différents selon le contexte des cytosines. Les contextes CpG et CpHpG sont appelés contextes symétriques, c'est-à-dire que lors de la réplication, les deux brins parentaux portent une trace de la méthylation. A l'inverse, le contexte CpHpH est asymétrique, la méthylation n'est portée que par un des brins parentaux.

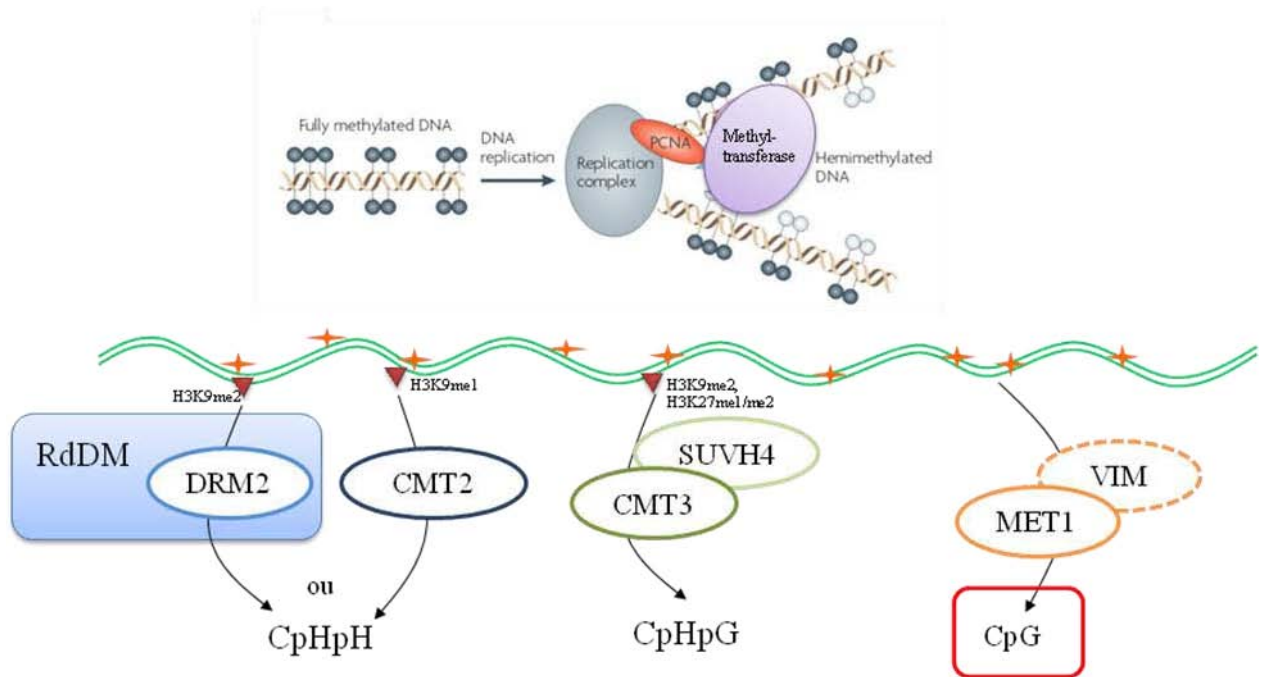


Figure 11 : Maintient de la méthylation de l'ADN au cours de la réplication

Adapté de (Law and Jacobsen, 2010) & (Stroud et al., 2014)

La méthyltransférase interagit avec le complexe de réplication au niveau de la fourche de réplication pour permettre le maintien de la méthylation des cytosines.

Les voies de maintien sont différentes selon le contexte de la cytosine. En contexte CpHpH, le maintien est assuré par DRM2 via le RdDM et par CHROMOMETHYLASE 2 (CMT2). En contexte CpHpG, le maintien se fait par CHROMOMETHYLASE 3 (CMT3). Enfin en contexte CpG, le maintien est assuré par METHYLTRANSFERASE 1 (MET1) avec l'aide du cofacteur VARIANT IN METHYLATION (VIM).

3.3.2.2. Maintien de la méthylation en contexte asymétrique CpHpH

Le maintien des méthylations en contexte asymétrique CpHpH est assuré par deux voies (Fig 11). La première est la voie RdDM *via* l'action de DRM2. Le double mutant d'*Arabidopsis drm2drm1* (*drm1* n'intervenant que dans les gamètes femelles) montre une nette diminution de la méthylation en contexte CpHpH, mais n'est pas suffisant pour supprimer la totalité des méthylations. Une deuxième méthyltransferase est impliquée dans le maintien des méthylations de type CpHpH : la CHROMOMETHYLASE2 (CMT2) (Stroud et *al.*, 2014). En effet, le triple mutant *drm1drm2cmt2* élimine la totalité des méthylations en contexte CpHpH. CMT2 est capable d'interagir directement avec les queues méthylées des histones H3, mais montre cependant une préférence de reconnaissance de la marque épigénétique H3K9me2. Il semble que la distribution des marques H3K9me1 et H3K9me2 détermine la fixation de l'une ou de l'autre des deux méthyltransferases. Les sites ciblés par CMT2 montrent un enrichissement en H3K9me2 comparés au site préférentiellement ciblé par DRM2, qui eux montrent un enrichissement en H3K9me1 (Stroud et *al.*, 2014).

3.3.2.3. Maintien de la méthylation en contexte symétrique CpG et CpHpG

Le gène *CHROMOMETHYLASE 3* (*CMT3*), a été identifié comme étant le principal acteur dans le maintien des méthylations en contexte CpHpG (Lindroth et *al.*, 2001) (Fig 11). Le mutant *cmt3-7* montre une perte quasi-totale sur l'ensemble du génome des méthylations CpHpG. Ceci est démontré par une analyse Southern Blot après digestion de l'ADN génomique par des enzymes sensibles à la méthylation révélée avec des sondes contre le rétrotransposon *Athila* (position péricentromérique) et les séquences satellites de 180pb (position centromérique) : *CMT3* n'a aucune action sur les méthylations en contexte CpG (Lindroth et *al.*, 2001). Le maintien de la méthylation par *CMT3* se ferait en association avec *SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE 4* (*SUVH4*). En effet, *SUVH4* possède un site de reconnaissance pour la méthylation CpHpG et est donc recruté par ces dernières. *SUVH4* est la principale histone méthyltransferase, mais possède deux paralogues impliqués dans le maintien des méthylations (Ebbs and Bender, 2006 ; Ebbs et *al.*, 2005). De même, *CMT3* possède un domaine protéique reconnaissant les marques H3K9 et H3K27 méthylées. Des études complémentaires soit d'immunoprécipitation de la chromatine (ChIP) soit d'hybridation *in situ*



Figure 12 : Représentation schématique du gène *MET1* chez le riz : *OsMET1a*

Les exons sont symbolisés par des rectangles et numéroté de 1 à 12. Les domaines protéiques sont symbolisés par des régions plus ou moins grisées et sont indiqués par les traits sur le dessus de la figure. Le domaine DNA methyltransferase est le domaine catalytique responsable de la méthylation de l'ADN.

au niveau des chromocentres montrent une coimmunoprécipitation / colocalisation entre les marques épigénétiques H3K9me2, H3K27me1 et H3K27me2 et la protéine CMT3 (Law and Jacobsen, 2010 ; Lindroth et *al.*, 2004). Le maintien en contexte CpG est quant à lui, assuré par le gène *METHYLTRANSFERASE 1 (MET1)* (Fig 10) dont le mode d'action est détaillé dans paragraphe 3.4.

3.3.2.4. Implication du gène *DECREASE IN DNA METHYLATION (DDM1)* dans le maintien de la méthylation

Le gène *DECREASE IN DNA METHYLATION (DDM1)* est également important pour le maintien des méthylations de l'ADN. Chez *Arabidopsis*, les mutants *ddm1-1* et *ddm1-2* montrent une diminution de 70% de la méthylation, dans tous les contextes de séquences (Vongs et *al.*, 1993). Il a très vite été montré que DDM1 n'était pas une méthyltransférase et que les mutants *ddm1* n'affectaient pas directement les méthyltransférases (Kakutani et *al.*, 1995). En effet, lorsque la séquence de *DDM1* a été obtenue, les études de phylogénie ont montré que cette protéine appartenait à la famille SWI2/SNF2, et plus particulièrement à une sous famille possédant des fonctions de remodelage de la chromatine (Jeddeloh et *al.*, 1999). L'action de *DDM1* s'exerce essentiellement au niveau des ETs (Lippman et *al.*, 2004), avec une préférence pour les ETs présents au niveau des centromères, que les auteurs ont définis comme des ETs hétérochromatiques par opposition aux ETs euchromatiques, présents sur les bras chromosomiques (Zemach et *al.*, 2013). L'action de DDM1 est dirigée par l'histone H1, qui se fixe au nucléosome et condense la chromatine. En effet, les plantes mutées pour l'histone H1 (*h1*) présentent une augmentation de la méthylation des ETs hétérochromatiques suggérant que l'histone H1 empêche l'accès des méthyltransférases à l'hétérochromatine. Enfin, le double mutant *h1ddm1* rétablit en partie le niveau de méthylation normale des ET hétérochromatiques (Zemach et *al.*, 2013).

3.4. Le gène *METHYLTRANSFERASE 1*

3.4.1. Structure du gène *MET1*

Chez *Arabidopsis*, *AtMET1* est un gène de 6 140 pb, composé de 11 exons et codant pour une protéine de 1534 acides aminés (aa) (P34881). Le riz compte deux gènes *MET1*, *OsMET1a* et *OsMET1b*. Il s'agit de gènes comptant 12 exons (Fig 12) codant pour une protéine de 1 527 aa pour *OsMET1a* (Yamauchi et *al.*, 2008). Chez *Brassica rapa*, il existe également

Arabidopsis

Nom du mutant	Méthode de génération des mutants	Références
met1-1	EMS	(Kankel et <i>al.</i> , 2003)
met1-2	EMS	
met1-3	T-DNA	(Saze et <i>al.</i> , 2003)
met1-4	T-DNA	
met1-5	EMS	(Boutet et <i>al.</i> , 2003)
met1-6	EMS	
met1-7	T-DNA	(Rocha et <i>al.</i> , 2005)
met1-h1	EMS	(Naumann et <i>al.</i> , 2005)
rts2-1	T-DNA	(Aufsatz et <i>al.</i> , 2004)
rts2-2	T-DNA	
Asmet1	construction anti-sens	(Finnegan et <i>al.</i> , 1996)
Asmet1	construction anti-sens	(Ronemus et <i>al.</i> , 1996)
DME:MET1 a/s	construction anti-sens	(Kim et <i>al.</i> , 2008)

Riz

Nom du mutant	Méthode de génération des mutants	Références
AsMET1 Nt	construction antisens	(Teerawanichpan et <i>al.</i> , 2004)
AsMET1 Ct	construction antisens	
pJHY-Ki-MET1a	mutant d'insertion : stratégie knock-in	(Yamauchi et <i>al.</i> , 2009)
pJHY-Ki-MET1b	mutant d'insertion : stratégie knock-in	(Yamauchi et <i>al.</i> , 2014)

Tableau 1 : Liste non exhaustive de mutants touchant le gène *MET1* chez les plantes

Liste de quelques mutants *met1* chez *Arabidopsis* et le riz, obtenus par différentes stratégies de validation fonctionnelle.

2 copies de *MET1* qui ont 10 et 11 exons et codent pour des protéines de 1 528 et 1 519 aa (Fujimoto et al., 2006). Chez le maïs il existerait un seul gène *MET1*, *ZmMET1*, codant pour une protéine de 1 525 aa (Steward et al., 2000). Les protéines MET1 sont caractérisées entre autre par la présence de trois types de domaines (Fig 9) : le domaine catalytique en partie C-terminale, le domaine méthylase qui caractérise la famille des DNA méthyltransférases. Ce domaine assure le transfert vers l'ADN d'un groupe méthyl apporté par la S-adenosyl méthionine. Deux domaines BAH (Bromo-Adjacent Homology) assurent les interactions protéine-protéine. Enfin, deux domaines DNMT1-RFD (DNMT1 Replication Foci Domain) dirige l'enzyme vers les foyers de réplication (site : pfam.sanger.ac.uk).

3.4.2. *MET1* et maintien des méthylations CpG

Il existe plusieurs mutants du gène *MET1* chez *Arabidopsis* et le riz (Tableau 1). Les premières publications décrivant *MET1* ont été publiées en 1996. Elles montrent une réduction assez importante de la méthylation de l'ADN dans des lignées exprimant un ARN antisens dirigé contre le gène *MET1*: pour certaines de ces lignées, la réduction de la méthylation est de 70-80% (Finnegan et al., 1996 ; Ronemus et al., 1996). *MET1* est impliqué dans le maintien des méthylations en contexte CpG. En effet les mutants *met1-1* et *met1-2* (fond génétique Col) montrent une réduction de 70% et 50% respectivement, de la méthylation des cytosines en contexte CpG par rapport au sauvage (Kankel et al., 2003). De même, dans le mutant *met1-3* (fond génétique Col), la méthylation au niveau des répétitions centromériques de 180pb est totalement éliminée en contexte CpG (Saze et al., 2003). *MET1* intervient aussi dans le maintien des méthylations en contextes CpHpG et CpHpH. Dans le mutant *met1-3* la méthylation en contexte CpHpG et CpHpH est réduite respectivement à 58% et 73% par rapport au niveau de méthylation d'une plante sauvage (Saze et al., 2003).

D'autres mutants *met1* obtenus dans l'écotype WS, montrent une diminution de la méthylation dans tous les contextes de méthylation pour 3 loci du gène *PAI*. Le niveau de méthylation en contexte CpG, CpHpG et CpHpH est réduit respectivement à 26%, 41% et 48% par rapport à une plante sauvage (Bartee and Bender, 2001). Enfin, l'analyse de la méthylation d'un SINE appelé *Sadhu* dans le mutant *met1-1* montre une réduction drastique de sa méthylation en contexte CpG mais qui s'accompagne aussi d'une diminution nette de la méthylation dans les deux autres contextes CpHpG et CpHpH (Rangwala and Richards, 2007). Même si par souci de simplification, *MET1* est le plus souvent décrit comme impliquée dans la méthylation CpG, la réalité n'est donc pas aussi simple. Il en va de même lorsque

méthylation *de novo* et méthylation de maintenance sont considérées. Le gène *MET1* joue un rôle prépondérant dans le maintien de la méthylation, mais il peut également intervenir lors de sa mise en place *de novo* plus classiquement attribuée à DRM2 (voir paragraphe 3.3.2.1.). Le mutant *rts2-2* (Tableau 1) est un mutant nul de *MET1* obtenu à partir d'une lignée préalablement transformée avec le promoteur NOS (NOSpro target) et l'élément silencer (NOSpro silencer), ce dernier étant nécessaire pour la méthylation du promoteur. Ce système permet l'étude de la mise en place et du maintien de la méthylation. Une forte diminution de la méthylation de maintenance a été décrite dans le mutant *rts2-2*, ainsi qu'une diminution par 3 de la méthylation *de novo* en contexte CpG, ce qui n'est pas le cas pour les autres contextes. *MET1* jouerait donc un rôle important dans la mise en place et le maintien de la méthylation en contexte CpG (Aufsatz et al., 2004).

3.4.3. Action de *MET1* pendant la réplication

Chez le riz et le maïs, l'expression de *MET1* est la plus forte dans les tissus en division, tels que les apex racinaires, les cals (tissus en prolifération issus de la culture cellulaire) ou dans les primordia des feuilles. A l'inverse, il n'est pas du tout exprimé dans les tissus différenciés ou les feuilles matures de riz. L'absence ou la faible expression dans les tissus différenciés est confortée par une étude chez le maïs où l'expression de *MET1* diminue à mesure que les tissus s'allongent ou se différencient (Steward et al., 2000 ; Teerawanichpan et al., 2004 ; Yamauchi et al., 2008). Chez *Arabidopsis* l'expression de *MET1* a été montrée dans l'œuf, dans les synergides et dans la cellule centrale, son expression étant toutefois également détectée dans l'embryon, l'endosperme et l'enveloppe de la graine mais à un moindre niveau (Schmitz et al., 2013).

L'expression d'un ARN antisens dirigé contre *MET1* sous contrôle du promoteur *DEMETTER*, l'une des déméthylases, exprimée dans le pistil mais également dans les cellules indifférenciées en prolifération de la jeune plantule, entraîne une hypométhylation des CpG dans toute la plante et ce de manière irréversible. L'inactivation de *MET1* au tout début du développement embryonnaire suffit donc à contrecarrer la mise en place de la méthylation et provoque des problèmes de développement pendant le reste de la vie de la plante (Kim et al., 2008).

La protéine MET1 possède deux domaines BAH et deux domaines DNMT1-RFD. Les domaines BAH sont connus pour être impliqués dans le ciblage de *DNMT1* au niveau des fourches de réplication chez la souris (Callebaut et al., 1999). Quant au second type de

domaine protéique, il est spécifique des méthyltransférases des cytosines et localise la protéine au niveau des foyers de réplication pendant la phase S par l'intermédiaire d'interaction avec d'autres partenaires protéiques (Rountree et *al.*, 2000). L'un des partenaires de DNMT1 chez les mammifères est l'Ubiquitin-like containing PHD and Ring Finger domain 1 (UHRF1), qui est capable d'interagir avec l'ADN hémiméthylé uniquement en contexte CpG et qui est nécessaire pour le maintien des méthylations en contexte CpG (Bostick et *al.*, 2007). Les homologues d'UHRF1 chez *Arabidopsis* sont les protéines VARIANT IN METHYLATION (VIM). Il existe cinq gènes *VIM* mais seuls *VIM1*, *VIM2*, *VIM3* sont exprimés. Le double mutant *vim1vim3* montre une forte diminution des méthylations en CpG dans les régions géniques et hétérochromatiques. Le triple mutant *vim1vim2vim3* montre des caractéristiques phénotypiques qui sont similaires à celles du mutant *met1*. Bien qu'il n'existe pas de preuve directe d'une interaction entre MET1 et VIM1/2/3, ces quatre protéines sont enrichies à l'interphase et colocalisent au niveau des chromocentres (Woo et *al.*, 2008). L'absence d'interaction directe suggère que d'autres partenaires seraient encore à découvrir.

3.4.4. *MET1* et les marques histones

Une autre conséquence de l'absence d'activité de MET1 est la redistribution d'autres marques épigénétiques sur la chromatine. Dans un mutant nul *met1-3* d'*Arabidopsis*, une diminution de la marque H3K9me2 est observée pour différentes cibles telles que le retrotransposon *Ta2*, les séquences du knob du chromosome 4 et les séquences centromériques de 180pb (Tariq et *al.*, 2003). Ces cibles, non transcrites dans un contexte sauvage, montrent une réactivation transcriptionnelle dans le contexte *met1* qui est corrélée avec une augmentation des marques actives H3K4me2/3. De façon assez surprenante, le niveau global de marque répressive H3K9me2 est le même chez *met1* que chez le sauvage. Toutefois, l'étude de la localisation par immunocytochimie sur des noyaux des marques 5mC, H3K9me2 et H3K4me2 montre que H3K9me2 et 5mC ne colocalisent plus avec les chromocentres comme c'est le cas chez la plante sauvage. Ceci prouve que la diméthylation H3K9 est redistribuée hors des chromocentres dans le mutant *met1*. Ainsi, la méthylation en CpG déterminerait la mise en place des marques H3K9me2 au niveau des chromocentres (Tariq et *al.*, 2003 ; Mathieu et *al.*, 2007).

3.4.5. *MET1* et la répression transcriptionnelle des éléments transposables

Comme présenté au paragraphe 3.2.1., la méthylation est essentielle pour réprimer les ETs. Cette répression n'est pas assurée par la seule protéine MET1, elle nécessite également l'HISTONE DEACETYLASE6 (HDA6). En effet, les protéines HDA6 et MET1 interagissent par l'intermédiaire de la région C-terminale d'HDA6 et du second domaine BAH de MET1. Cette interaction est nécessaire à la régulation de la méthylation de l'ADN, de la méthylation et de l'acétylation des histones dans le but de maintenir les ETs réprimés. Certains de ces ET réactivés dans des mutants *hda6* montrent une augmentation de l'acétylation des histones H3 et H4, accompagnée d'une augmentation des marques H3K4me3 et H3K4me2, ainsi qu'une diminution importante de la méthylation dans les trois contextes (Liu et al., 2012).

3.4.6. *MET1* et le développement

Chez *Arabidopsis*, en plus des « phénotypes moléculaires », les mutants *met1* montrent plusieurs anomalies de développement causées par les changements d'expression du génome consécutifs à la diminution de la méthylation. L'une des plus marquantes est le retard de floraison. Le développement anormal des structures végétatives, la diminution de la fertilité et une diminution de la dominance apicale sont également observés. Il semble que plus la diminution de la méthylation est importante plus les phénotypes sont sévères (Finnegan et al., 1996). Les protéines du groupe *Polycomb* (*Polycomb-group*, *PcG*) que sont MEDEA (MEA), FERTILIZATION INDEPENDENT ENDOSPERM (FIE) et FERTILIZATION INDEPENDENT SEED2 (FIS2) sont nécessaires au contrôle du développement de la graine chez *Arabidopsis*. Les mutants *mea* présentent un développement aberrant des embryons ainsi qu'une mort prématurée (Grossniklaus et al., 1998), alors que les mutants *fis* et *fie* montrent un développement de la graine sans nécessité de fécondation (Ohad et al., 1996 ; Chaudhury et al., 1997). Ces trois gènes sont soumis à l'empreinte parentale, avec une expression maternelle (Grossniklaus et al., 2001), leur activité étant liée à celle de MET1. Tout d'abord, des analyses de double hybride ont montré une interaction de MET1 avec MEA, ainsi qu'avec FIS, *via* les domaines BAH et la région N-terminale. Ensuite, les gènes *MET1*, *MEA* et *FIS* montrent des profils d'expression similaires. Enfin, des plantes *mea-1/MEA met1-3/MET1* montrent une fréquence plus élevée d'ovules avec plus de deux noyaux dans la cellule centrale, et une fréquence plus élevée de développement autonome de l'endosperme. MET1 et MEA agissent de façon synergiques dans

la répression du développement de l'endosperme sans fertilisation (Schmidt et *al.*, 2013).

MET1 est un gène important pour le développement de la plante et ceci se confirme chez le riz. En effet, les mutants nuls *met1* sont viables chez *Arabidopsis*, mais ce n'est pas le cas chez le riz. La copie du riz *OsMET1b*, s'exprime jusqu'à 12 fois plus que la seconde copie, *OsMET1a*. Le mutant *osmet1a* est viable alors que ce n'est pas le cas du mutant *osmet1b* (Yamauchi et *al.*, 2014).

Objectifs

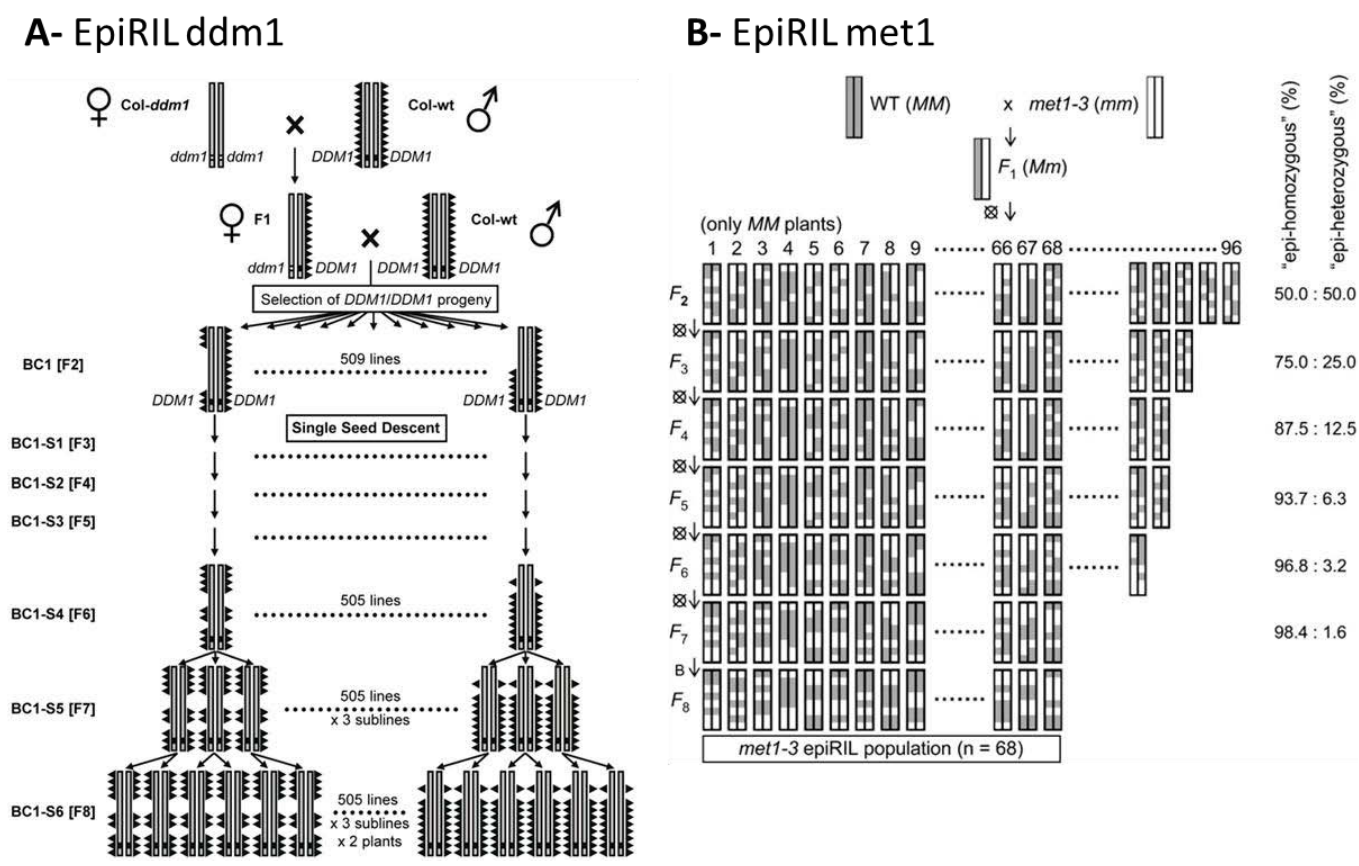


Figure 13 : Etablissement des EpiRIL *met1* et *ddm1* chez *Arabidopsis thaliana*

Source : (Johannes et al., 2009a) & (Reinders et al., 2009a)

A- Le génome d'*Arabidopsis* est représenté en gris. Les triangles symbolisent la présence de méthylation sur l'ADN. La localisation de *DDM1* est symbolisée par des carrés noirs pour la version sauvage et blancs pour la version mutée. Après un premier back-cross (BC1), les lignées sont autofécondées (S1-S6) jusqu'à la huitième génération (F8).

B- Les chromosomes et segments de chromosomes sauvages sont en gris, et ceux correspondant au mutant *met1-3* en blanc. Les huit générations d'autofécondation sont indiquées de F2 à F8. Les pourcentages d'epi-homozygotie ou epi-hétérozygotie prédits sont indiqués sur la droite.

C'est certainement la découverte récente chez *Arabidopsis* de la possibilité de créer des populations de type EpiRILs, c'est-à-dire des populations présentant une altération stable de leur profil de méthylation (Johannes et al., 2009; Reinders et al., 2009; Richards, 2009) qui a été l'un des éléments déclencheur de mon projet de thèse. Ces populations sont établies par croisement d'un mutant homozygote pour la mutation *met1* ou *ddm1*, possédant donc un génome hypométhylé, avec une plante de génotype sauvage. La version homozygote sauvage du gène *MET1* ou *DDM1* est ensuite rétablie dès la seconde génération du croisement (Fig 13). Après autofécondation sur plusieurs générations, les profils de méthylation qui dans un premier temps sont en mosaïques, sont peu à peu remis en place. Après huit générations, les profils de méthylation sont rétablis à hauteur de 80% du sauvage. Il existe cependant des « accidents » de restauration du profil de méthylation qui sont hérités de façon stable sur plusieurs générations. Il y a donc création d'épi-allèles. La création des lignées EpiRILs s'accompagne de variations phénotypiques pour des caractères tels que le temps de floraison ou la taille des plantes. Toutefois, il faut noter que certains éléments transposables semblent également déréprimés au cours de l'établissement des lignées. Ainsi, le rétrotransposon EVD (voir chapitre introduction, (Mirouze et al., 2009)) ou d'autres éléments de type CACTA (Johannes et al., 2009; Reinders et al., 2009) s'accumulent au cours des générations. Ces données présentent un intérêt si elles s'avèrent transposables chez le blé. En effet, ces résultats suggèrent que la variabilité épigénétique peut modifier le phénotype d'une lignée et que cette modification est héritable. Il serait ainsi possible à long terme d'utiliser la variabilité épigénétique au même titre que la variabilité génétique dans l'amélioration des variétés.

Ces données ayant été obtenues avec des mutants *met1* et *ddm1*, qui sont bien caractérisés chez *Arabidopsis*, mon projet de thèse a consisté à analyser la fonction de ces deux gènes mais cette fois-ci chez le blé. Mon projet de thèse s'est déroulé dans le cadre d'une convention Cifre, soit un partenariat public/privé entre la société de biotechnologies végétales Biogemma et le laboratoire GReD (Génétique, Reproduction et Développement) qui est une Unité Mixte de recherche (UMR CNRS 6293, INSERM U1103 GReD, Clermont Université). Initialement, mon projet s'est divisé en trois parties : (i) la création de lignées de blé modifiées pour leur épigénome, (ii) la mise au point d'une méthode de suivi des profils de méthylation chez le blé et enfin (iii) la caractérisation du profil de méthylation des lignées modifiées pour une sélection de gènes (approche gène candidat) (Fig 14). Comme mon travail est centré sur les gènes *MET1* et *DDM1*, non connus chez le blé, le premier point était d'identifier les séquences

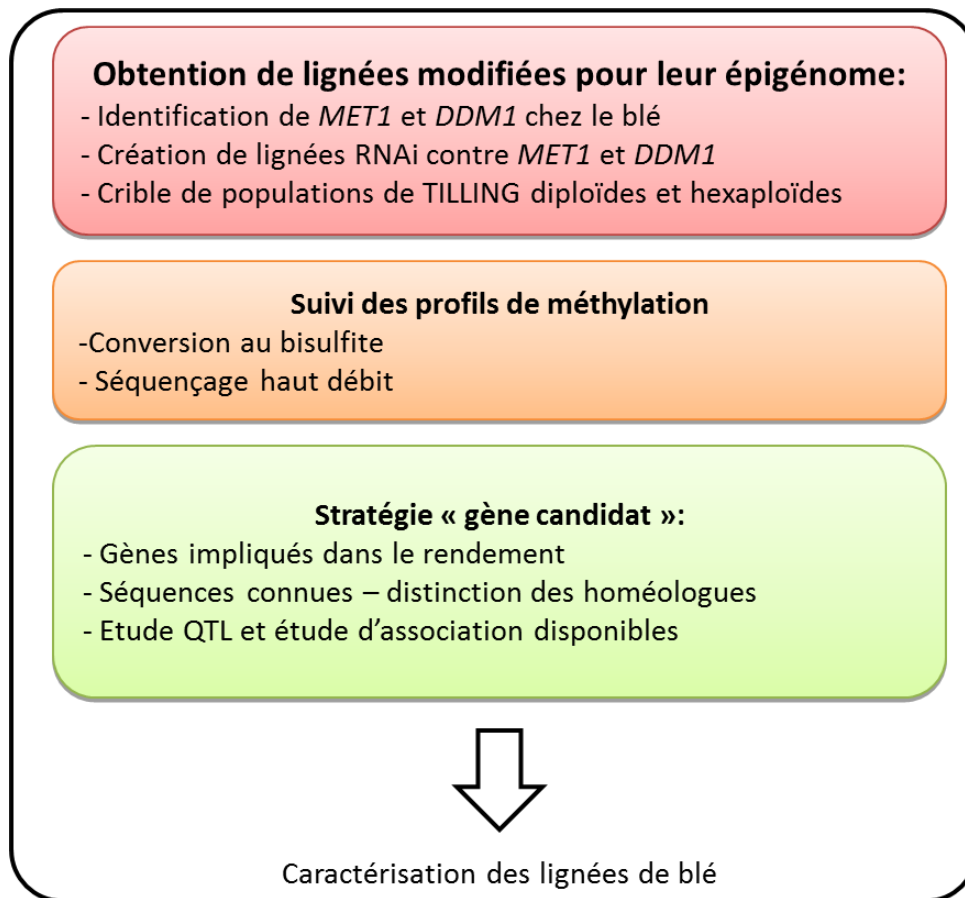


Figure 14 : Principales phases du projet de thèse

de ces gènes. Ce travail est l'objet du premier chapitre de ma thèse. Une fois identifiés, ces gènes seront utilisés afin d'obtenir des lignées modifiées pour leur profils de méthylation. Différentes stratégies ont été envisagées, à la fois sur le blé diploïde et le blé hexaploïde. Des mutants seront recherchés dans des populations de mutagènes dites de TILLING (McCallum et *al.*, 2000), de telles ressources sont en effet disponibles chez le blé diploïde et hexaploïde. Nous avons également envisagé la création de lignées transgéniques capables de réprimer l'expression des deux gènes d'intérêt *via* une approche d'interférence ARN (RNAi) (Fu et *al.*, 2007). Ces deux approches seront décrites dans le chapitre 2. Un autre volet de mon travail de thèse consiste à caractériser les lignées modifiées sur leur profil de méthylation, également décrit dans le chapitre 2. Ce travail nécessite d'établir une méthode permettant d'estimer l'état de méthylation des cytosines dans les contextes CpG, CpHpG et CpHpH. Cette méthode sera ensuite utilisée pour caractériser les lignées modifiées. Parmi l'ensemble des méthodes disponibles au démarrage du projet (Laird, 2010), nous avons retenu celle de la conversion au bisulfite, capable de modifier de façon spécifique les cytosines non méthylées. Cette méthode sera suivie d'un séquençage haut débit de type 454 (Margulies et *al.*, 2005), technique disponible à Biogemma. Enfin, comme nous ne souhaitons pas réaliser d'emblée une approche sur la totalité du génome, il a été fait le choix d'une stratégie gènes candidats. Ces gènes devront être définis selon différents critères tels que leur intérêt sur le plan agronomique, en particulier d'un point de vue du rendement, et le fait que leurs séquences soient connues chez le blé hexaploïde.

L'ensemble de ces trois volets est résumé dans la figure 14.

Résultats

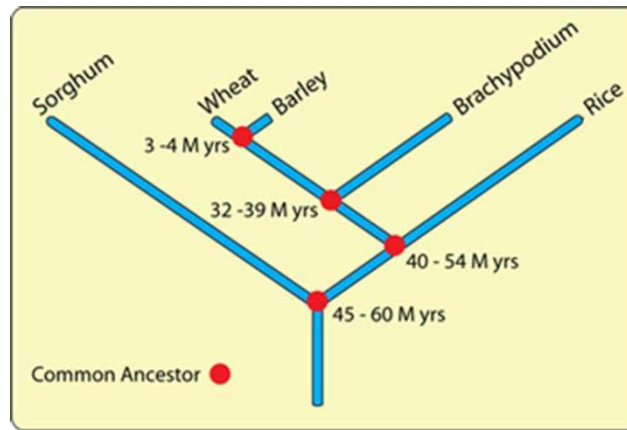


Figure 15: Arbre phylogénétique retraçant l’histoire évolutive des espèces ayant un ancêtre commun avec le blé

Source : http://www.wheatbp.net/WheatBP/Images/Evolution_Dendrogram_small.png

Cet arbre permet de visualiser la spéciation du blé et de l’orge (tribu des *Triticeae*) qui serait survenue il y a 3-4 Millions d’années (Myrs). Les analyses de comparaison des génomes (synténie) utilisent souvent le riz et *Brachypodium*, deux espèces ayant divergées il y a respectivement 40-54 et 32-39 Myrs par rapport au blé. Le sorgho est également indiqué sur cet arbre bien qu’il soit en général utilisé plutôt lors de comparaison avec le maïs.

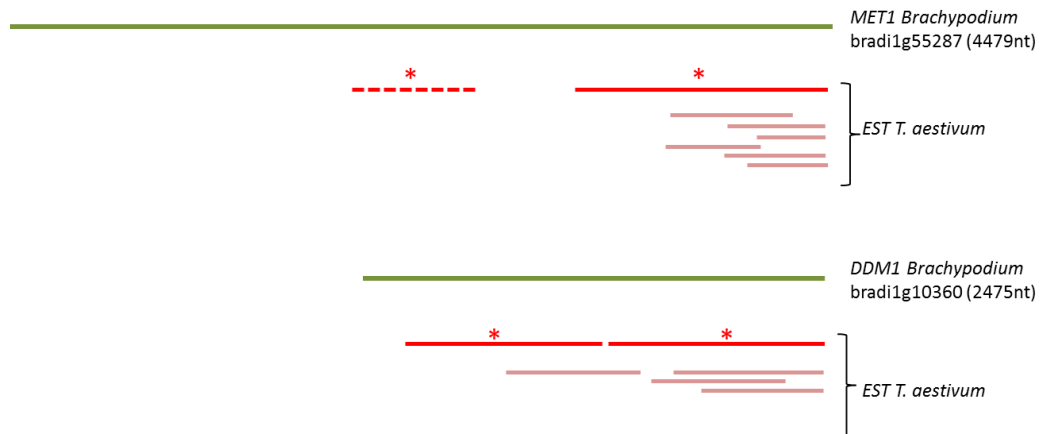


Figure 16: Représentation schématique des ESTs des gènes *MET1* et *DDM1* identifiées chez le blé tendre

La localisation a été faite par une analyse BLASTn entre les ESTs de blé (en rouge et rose) et les séquences codantes (CDS) de *Brachypodium* (en vert). Les traits pleins correspondent aux ESTs publiques et le trait pointillé à l’EST *MET1* trouvée dans les bases de données privées de Biogemma. La recherche de similarité par BLAST a permis d’identifier 8 ESTs correspondant à *MET1* : 7 publiques : TA83026_4565, CA499915, CJ576093, BJ309923, CA639972, TA83027_4565, CA645113 (NCBI UniGene) et une privée : GPIC:S:720428. Pour *DDM1*, 6 ESTs ont été identifiées : TA50412_4565, TA93691_4565, TA1932_4565, CA499811, CV761210, CK218180.

Les EST indiquées par une astérisque (*) ont été utilisées pour définir les oligonucléotides lors de la première expérience de Sequence Capture : les contigs TA83026_4565, GPIC:S:720428 pour *MET1* et TA93691_4565, TA50412_4565 pour *DDM1*.

Chapitre 1 : Les gènes *TaMET-1* chez le blé tendre

1. Historique de la découverte des gènes *MET1* chez le blé tendre

Lorsque que le travail sur *MET1* et *DDMI* a débuté, aucune des séquences de ces deux gènes n'était connue chez le blé. Début 2011, le génome du blé n'était pas séquencé et les seules données de séquences disponibles chez Biogemma étaient des données d'expression (Expressed Sequence Tag, EST) et des données de séquençage effectuées sur la variété *Chinese Spring* par le groupe Functional Genomics de l'université de Bristol (Brenchley et al., 2012). Il s'agit d'un séquençage généré par la méthode 454 de Roche, présentant une couverture limitée de séquençage de l'ordre de 5X ce qui suggérait d'après les indications du site internet (<http://www.cerealsdb.uk.net>) un certain taux d'erreurs de séquençage non corrigées. Pour nous aider, nous disposions également des séquences de *MET1* et *DDMI* connues chez d'autres espèces telles qu'*Arabidopsis* (dicotylédone) et le riz (monocotylédone) par exemple, ce dernier compte deux copies (deux paralogues) de *MET1*, appelée *OsMET1a* et *OsMET1b*, et deux de *DDMI*. Cette première observation de l'existence d'une duplication de gènes chez les monocotylédones sera d'ailleurs une information importante dans les phases initiales de mon projet. Dans les phases préliminaires du travail, j'ai pu me rendre compte de l'utilité des espèces évolutivement proches du blé (Fig 15). La séquence complète du génome d'*Arabidopsis* est disponible depuis 2000 (The Arabidopsis Genome Initiative, 2000), celle du riz depuis 2005 (International Rice Genome Sequencing, 2005), celle de *Brachypodium* depuis 2010 (Vogel et al., 2010) et celle de l'orge l'a été au cours de notre projet en 2012 (The International Barley Genome Sequencing, 2012). Les données du riz, de *Brachypodium* et de l'orge m'ont par la suite aidé à mieux comprendre la structure des gènes, l'épissage et la structure des ARNm ainsi que leurs protéines. Une analyse BLASTn sur des bases de données des régions codantes (CDS) de *Brachypodium* à partir des séquences d'*Arabidopsis* et du riz a permis d'identifier les séquences de nos gènes, par exemple pour *MET1* : bradi1g55287 (gène *MET1a*) et bradi1g05380 (gène *MET1b*). Ces premières



Figure 17: Découverte de nouvelles séquences génomiques chez le blé par Sequence Capture

Les EST de blé et les séquences génomiques de *Brachypodium* ont été utilisées pour définir les sondes oligonucléotidiques sur la puce ADN (micro array). L'ADN génomique des variétés *Chinese Spring*, *Renan*, *Recital*, *Brigadier* et *Alcedo* ont été utilisés afin de capturer les séquences des gènes *MET1* et *DDM1*. L'expérience a été réalisée chez Biogemma et supervisée par Jean-Philippe Pichon. L'assemblage des séquences a été fait à Biogemma par Jorge Duarte. Cette expérience a également permis d'identifier des SNPs entre les parents de deux populations de cartographie utilisées à Biogemma: entre les variétés *Renan* et *Recital* (population de cartographie ReR) d'une part et les variétés *Brigadier* et *Alcedo* (population de cartographie Bal) d'autre part. J'ai réalisé le travail de détection des SNPs avec l'aide de Sébastien Faure de Biogemma.

informations collectées chez le riz et/ou *Brachypodium* nous ont fait supposer qu'il y aurait au moins deux paralogues de *MET1* et *DDMI* pour chaque génome homéologue du blé tendre. En d'autres termes, nous nous attendions à découvrir chez le blé tendre au moins 6 copies de chacun de nos deux gènes candidats.

Une recherche de similarité de séquences contre l'ensemble des données publiques et privées disponibles à Biogemma a été réalisée en utilisant les séquences de *MET1* et *DDMI* identifiées chez ces trois espèces (*Arabidopsis*, riz et *Brachypodium*). Plusieurs contigs d'EST ont été identifiés, après vérification de leur pertinence et spécificité, seulement 8 contigs ont été retenus comme correspondant au gène *MET1* et 6 contigs correspondant à *DDMI* chez le blé tendre (Fig 16). Si pour *DDMI*, les ESTs couvraient bien le gène, pour *MET1*, les ESTs étaient toutes dans la région 3' et ne couvraient pas la totalité de la séquence. Il faut en effet préciser ici que le gène *MET1* est un gène d'assez grande taille (5844nt et 6035nt pour les deux gènes *MET1* chez *Brachypodium*). Il est également cohérent que toutes ces ESTs correspondent à la partie 3' du gène, car elles ont été obtenues par transcription inverse à partir de la queue PolyA des ARNm. Il est toutefois surprenant qu'au vu de la taille de la banque d'EST, seulement huit séquences pour *MET1* et six pour *DDMI* aient pu être identifiées. Cela suggérerait une expression potentiellement faible, ou une expression dans des tissus mal représentés dans les banques d'ESTs. Il a donc fallu définir une stratégie afin de connaître la séquence totale de nos gènes *MET1* et *DDMI*. Nous avons profité du savoir-faire de Biogemma pour certaines méthodes haut débit couramment utilisées en génomique. Ainsi, la technologie des microarrays (puce à ADN composée d'oligonucléotides de séquence connue) permet de capturer des séquences génomiques s'hybridant avec les oligonucléotides immobilisés sur la puce. Ces séquences sont ensuite éluées puis soumises à un séquençage nouvelle génération (NGS). Cette stratégie dite du « Sequence Capture » avait déjà été utilisée pour obtenir le séquençage de région exonique de 3,5Mb couvrant près 3 500 gènes chez le blé (Saintenac et al., 2011a) et est couramment utilisée par Biogemma (Fig 17).

La stratégie mise en place pour *MET1* pour le Sequence Capture est décrite dans la partie Matériels et Méthodes de notre article. La même stratégie a été utilisée pour *DDMI* mais avec des oligonucléotides différents pour capturer les séquences génomiques du gène *DDMI*. Deux vagues de Sequence Capture ont été réalisées (appelées W05 et W06 selon la nomenclature interne de Biogemma). La première vague, a été faite en utilisant essentiellement les séquences ESTs décrites en figure 13 puis avec les séquences issues des gènes identifiés par analyse BLAST chez *Brachypodium*. A la fin de ces deux vagues de Sequence Capture, peu de séquences

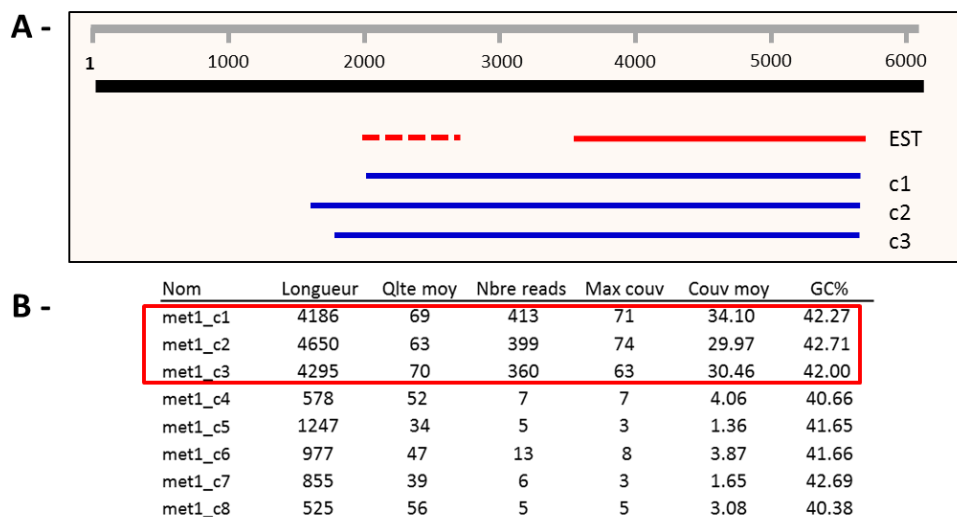


Figure 18 : Résultats de la première vague de Sequence Capture pour le gène *MET1* (expérience W05)

A- Schéma des régions couvertes par les ESTs et les contigs résultant du Sequence Capture sur la séquence génomique *Brachypodium*. La séquence bradi1g55287 (en noir), les deux ESTs de blé (en rouge) et les trois contigs Sequence Capture c1, c2 et c3 les plus longs (en bleu).

B- Données Sequence Capture pour les huit premiers contigs. Les données concernant les contigs c1, c2 et c3 sont encadrées en rouge.

A- Paramètres d'assemblage moins stringents : 3 contigs

Nom	Longueur	Qlte moy	Nbre de reads	Max couv	Couv moy	GC%
met1_rep_c32	6871	72	1608	185	87.03	42.77
met1_rep_c31	6475	72	1489	169	84.54	41.76
met1_rep_c34	6409	72	1452	175	85.58	42.00

B- Paramètres d'assemblage par défaut : 5 contigs

Nom	Longueur	Qlte moy	Nbre de reads	Max couv	Couv moy	GC%
met1_c9	5914	69	214	31	13.92	43.42
met1_c3	5876	70	256	35	15.71	43.64
met1_c5	2852	66	103	39	13.08	43.72
met1_c15	2536	66	55	20	8.11	41.93
met1_c1	2415	60	60	22	8.98	42.19

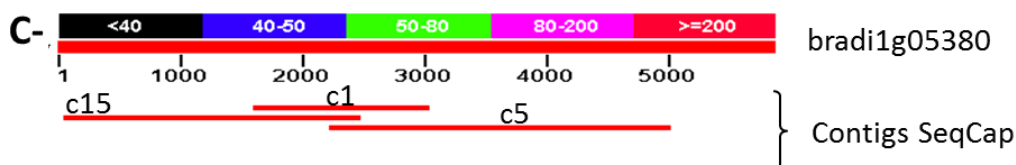


Figure 19: Résultats combinés des deux vagues de Sequence Capture (SeqCap) pour le gène *MET1* (expérience W05 + W06).

En fonction des paramètres d'assemblage, différents contigs ont pu être assemblés.

A- Avec des paramètres moins stringents, 3 contigs ont été reconstruits : met1_rep_c31/c32/c34. Ces trois contigs correspondent respectivement aux contigs c2, c1 et c3 de la figure 4.

B- Avec des paramètres par défaut, 5 contigs ont été obtenus, dont deux contigs de grandes tailles, met1_c3 et met1_c9. Parmi les contigs de plus petites tailles, met1_c15 et met1_c5 sont chevauchant et pourraient être assemblés en modifiant les paramètres de stringence. Le contig met1_c1 est également chevauchant mais présente un grand nombre de polymorphismes, ne permettant pas d'envisager un assemblage.

C- Résultats de l'alignement BLAST des contigs c1, c5 et c15 contre la séquence génomique de bradi1g05380.

exploitables avaient été identifiées pour le gène *DDMI*. Le travail sur ce gène a donc été stoppé (réunion d'avancement GReD/Biogemma Février 2012) et ne sera pas présenté dans ce mémoire. La suite de mon projet de thèse s'est focalisée sur *MET1*.

Pour le gène *MET1*, la première vague a permis de reconstruire 3 contigs couvrant la fin du gène, le début étant toujours manquant (Fig 18). La taille cumulée des deux ESTs est de 2500 bases et ils couvrent un peu plus de 2800 bases de la séquences génomique de *Brachypodium* en tenant compte des introns. Le Sequence Capture a tout d'abord permis de définir les séquences introniques, la région entre les deux ESTs, mais également d'obtenir plus de 350 bases en amont de l'EST le plus en 5'. Au total, cette première expérience a permis de découvrir 1500 pb supplémentaires dont la plupart des séquences se regroupent en trois principaux contigs (c1, c2 et c3 sur la Fig 18B). En plus des trois contigs principaux, d'autres contigs ont été générés mais semblent ne pas pouvoir s'assembler (Fig 18B).

En parallèle de la première vague de Sequence Capture, le gène *MET1* a été recherché par amplification PCR chez l'espèce diploïde *T. monococcum*. Cette espèce diploïde ($2n = 7$) est proche de l'ancêtre du génome A du blé tendre, *T. urartu* dont le génome a été séquencé récemment (Ling et al., 2013) mais n'était pas disponible au moment de ce travail. Notre hypothèse était que comme chez les autres céréales diploïdes (*Brachypodium* et riz), le génome de *T. monococcum* devrait également comporter deux paralogues de *MET1*. Pourtant une seule séquence a pu être amplifiée.

Bien qu'il ait été assez difficile de comprendre ces résultats nous avons fait l'hypothèse que les ESTs dont nous disposions au départ du projet pourrait dériver de la même copie de *MET1* et que les principaux contigs reconstitués par Sequence Capture correspondraient également à ces ESTs. Toutefois, la présence de contigs supplémentaires encore très incomplets et le fait que nous connaissions l'existence de deux copies *OsMET1a* et *OsMET1b* chez le riz et *Brachypodium* (bradi1g55287 et bradi1g05380) nous a conforté dans l'idée qu'il existerait aussi deux paralogues chez le blé avec chacun trois copies homéologues (génome A, B et D).

Une seconde vague de Sequence Capture a donc été lancée pour compléter les trois premiers contigs et en générer de nouveaux. La deuxième vague a été lancée en utilisant les trois contigs principaux c1, c2 et c3 ainsi que la séquence génomique *MET1a* de *Brachypodium* la plus éloignée des contigs de *MET1* déjà identifiés. A l'issue de cette nouvelle analyse, les trois premiers contigs ont été complétés et trois nouveaux ont été identifiés (Fig 19). Nous

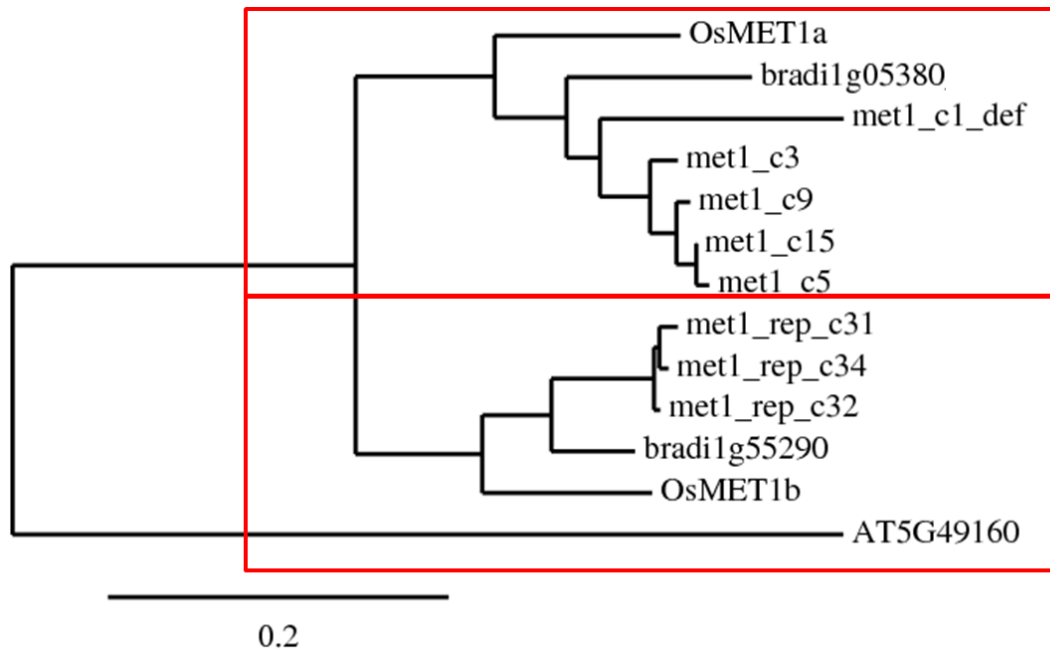


Figure 20: Analyse phylogénétique des contigs Sequence Capture

L'analyse inclue à la fois la séquence des espèces modèles (riz, *Brachypodium* et *Arabidopsis*) et les séquences de blé issues des deux vagues de Sequence Capture (W05 et W06) : les contigs utilisés sont ceux présentés à la figure 15 & 16. Ces premiers éléments nous ont permis d'émettre l'hypothèse qu'il pourrait exister 2 paralogues de *MET1* chez le blé. Toutefois quelques contigs ne se regroupent pas avec les principaux contigs suggérant qu'il pourrait exister plus que deux paralogues : contigs met1_c1.

Alignement MUSCLE (paramètres par défauts) puis construction de l'arbre avec PhyML et TreeDyn disponible sur <http://www.phylogeny.fr/>

avons donc six copies de *MET1* qui semblaient se répartir en deux grands groupes *MET1a* et *MET1b* d'après nos analyses phylogénétiques (Fig 19). Nous avons tout de même un doute sur le fait qu'il n'existait que deux paralogues et pour chacun d'eux trois homéologues, car certaines séquences ne s'assemblaient pas avec les six principaux contigs (Fig 20).

Un début de réponse est arrivé grâce à l'étude des quatre génotypes utilisés pour le Sequence Capture (voir Fig 17) qui ont permis de réaliser la cartographie des différentes copies de *MET1* (voir publication). Seuls les résultats de la seconde vague ont permis d'identifier des SNPs discriminant les parents des populations de cartographie. *MET1* a été positionné sur les chromosomes 7B (Bal), 7A (Bal, ReR), 5B (ReR), 5A (ReR, Bal), 2B (ReR). Ces résultats de cartographies semblent montrer la présence de *MET1* sur les groupes de chromosomes 2, 5 et 7. Il y aurait donc trois paralogues de *MET1* et pour chaque paralogue, trois copies homéologues soit au maximum neuf gènes ! Chacune des positions de cartographie ayant été obtenue avec plusieurs marqueurs, nous pensons qu'il était peu probable qu'il s'agisse d'erreur de cartographie. Enfin, les données du consortium de séquençage du génome du blé (IWGSC) sont arrivées en fin d'année 2012. Dans un premier temps seul Biogemma a eu accès aux données, nous avons ensuite obtenu un accès via l'URGI (INRA). En une seule et simple analyse BLASTn, j'ai pu confirmer la présence de neuf copies de *MET1*, sur les chromosomes 2, 5 et 7 et pour les trois génomes homéologues. Comme il s'agissait d'une première version du séquençage pouvant contenir des erreurs dans l'assemblage des données, ces résultats ont été confirmés par amplification PCR sur le diploïde *T. monoccocum* et l'hexaploïde.

En bilan de cette partie replaçant l'historique de mon projet, je dirai que l'étude par la méthode de capture de séquence a demandé un investissement important tant à Biogemma que de ma part. Bénéficier aujourd'hui de la ressource de séquence génomique du blé tendre est un gain de temps énorme pour la communauté travaillant sur le blé. Cette ressource nous aurait sans doute fait gagner un an sur mon projet de thèse.

2. Mise au point de la conversion au bisulfite

Avant de pouvoir faire l'étude des profils de méthylation chez le blé, il a fallu mettre au point la conversion au bisulfite suivie d'un séquençage haut débit. Cette méthode permet de convertir les cytosines non méthylées en uraciles alors que les cytosines méthylées restent inchangées. Par séquençage et comparaison avec un ADN non traité, il sera donc possible de déterminer le profil de méthylation d'une séquence. Ce traitement implique que chaque brin

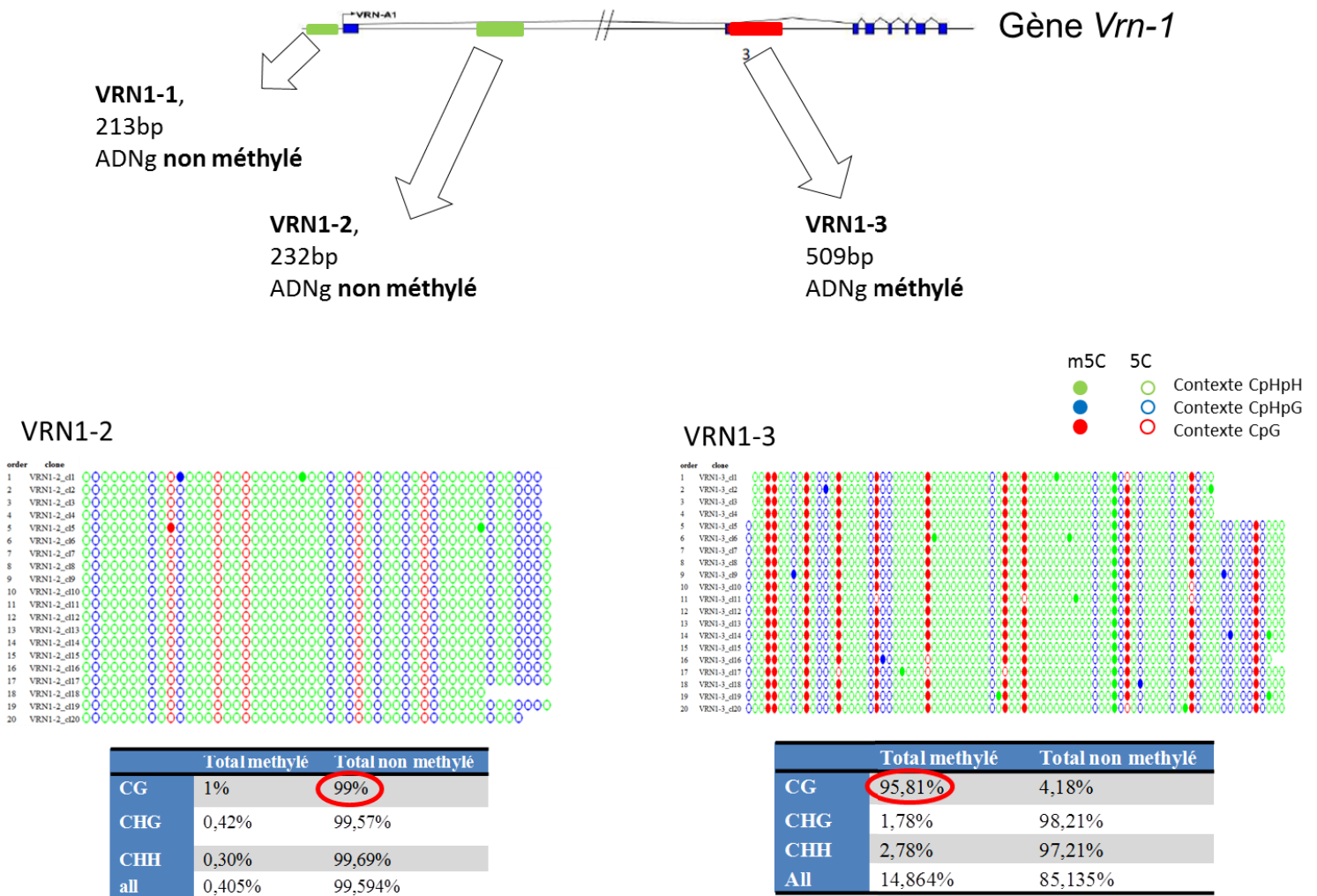


Figure 21 : Localisation des trois amplicons VRN1 utilisés pour la mise au point de la conversion au bisulfite

L'amplicon VRN1-1 est situé en amont de la région codante, l'amplicon VRN1-2 au niveau du premier intron et l'amplicon VRN1-3 au niveau du deuxième exon et intron.

Les trois amplicons ne présentent pas la même spécificité d'amplification. Ils peuvent être spécifiques d'un génome (VRN1-2 et VRN1-3) ou amplifier les trois génomes homéologues (VRN1-1). Ils présentent des profils de méthylation distincts: pas de méthylation (VRN1-1 et VRN1-2) ou forte méthylation en contexte CpG (VRN1-3).

Un exemple de profil bisulfite est donné pour VRN1-2 et VRN1-3 (sorties graphiques issues du logiciel Kismeth). Nos résultats ici obtenus après séquençage par la méthode de Sanger confirment ceux de l'équipe de J. Enjalbert et montrent une faible méthylation (cercles vides) dans les trois contextes pour VRN1-2 et une forte méthylation dans le contexte CpG pour l'amplicon VRN1-3 (cercles rouges pleins).

d'ADN soit converti en une nouvelle séquence qui sera distincte entre les deux brins. Lors des discussions initiales sur mon projet, nous avons fait le choix de n'étudier qu'un seul des deux brins convertis.

Cette mise au point a été faite en partenariat avec l'équipe de J. Enjalbert (INRA du Moulon), qui nous a fourni les séquences et les amorces permettant d'amplifier trois amplicons sur le gène VERNALIZATION 1 (*Vrn-1*) : VRN1-1, VRN1-2, VRN1-3 (Fig 21), ce dernier étant fortement méthylé en contexte CpG. Cette mise au point a permis de montrer que l'un des facteurs limitant était la quantité d'ADNg utilisé au départ de la conversion au bisulfite. L'ensemble de cette mise au point est disponible en annexes. Dans notre publication, l'étude des profils de méthylation s'est focalisée sur les régions en amont des gènes *TaMET-1* présents sur les chromosomes 5 et 7. En effet, un couple par génome homéologue avait été défini pour la région en amont des gènes du groupe 2 mais après analyse du séquençage, il s'est avéré que la région ciblée correspondait à un élément répété, non identifié malgré une recherche précédente réalisée sur les bases d'éléments répétés (TREP ; <http://wheat.pw.usda.gov/ITMI/Repeats/>). Pour une question de temps il n'a pas été possible d'intégrer l'analyse des régions promotrices du groupe 2 dans la version de notre publication. Cette analyse sera faite dans le cadre de la caractérisation des lignées modifiées pour leur épigénome (voir chapitre 2).

3. Préambule

La figure 22 récapitule notre démarche expérimentale. Dans notre publication nous nous sommes concentrés sur l'analyse de l'histoire évolutive de la famille *MET1* et à essayer de comprendre pourquoi le génome du blé tendre dispose de 9 copies de *TaMET-1*. Nous avons au cours de ce travail profité de l'apport d'Etienne Paux et Lise Pingault (INRA de Crouël) et de leurs données non publiées de RNA-Seq. Nous avons également bénéficié des conseils de Gisèle Bronner et Simon Roux (laboratoire LMGE, Clermont-Ferrand) pour l'ensemble des analyses phylogénétiques et de vitesse d'évolution. Enfin, nous avons proposé notre article une première fois à Genome Biology via Kellie Eversole, directeur exécutif de l'IWGSC. En effet Genome Biology souhaite faire un numéro spécial incluant des études comme la notre ayant largement bénéficiées du séquençage du génome du blé. Malheureusement l'article a été refusé et Genome Biology a suggéré sa soumission à BMC Genomics, ce qui sera fait début mai 2014.

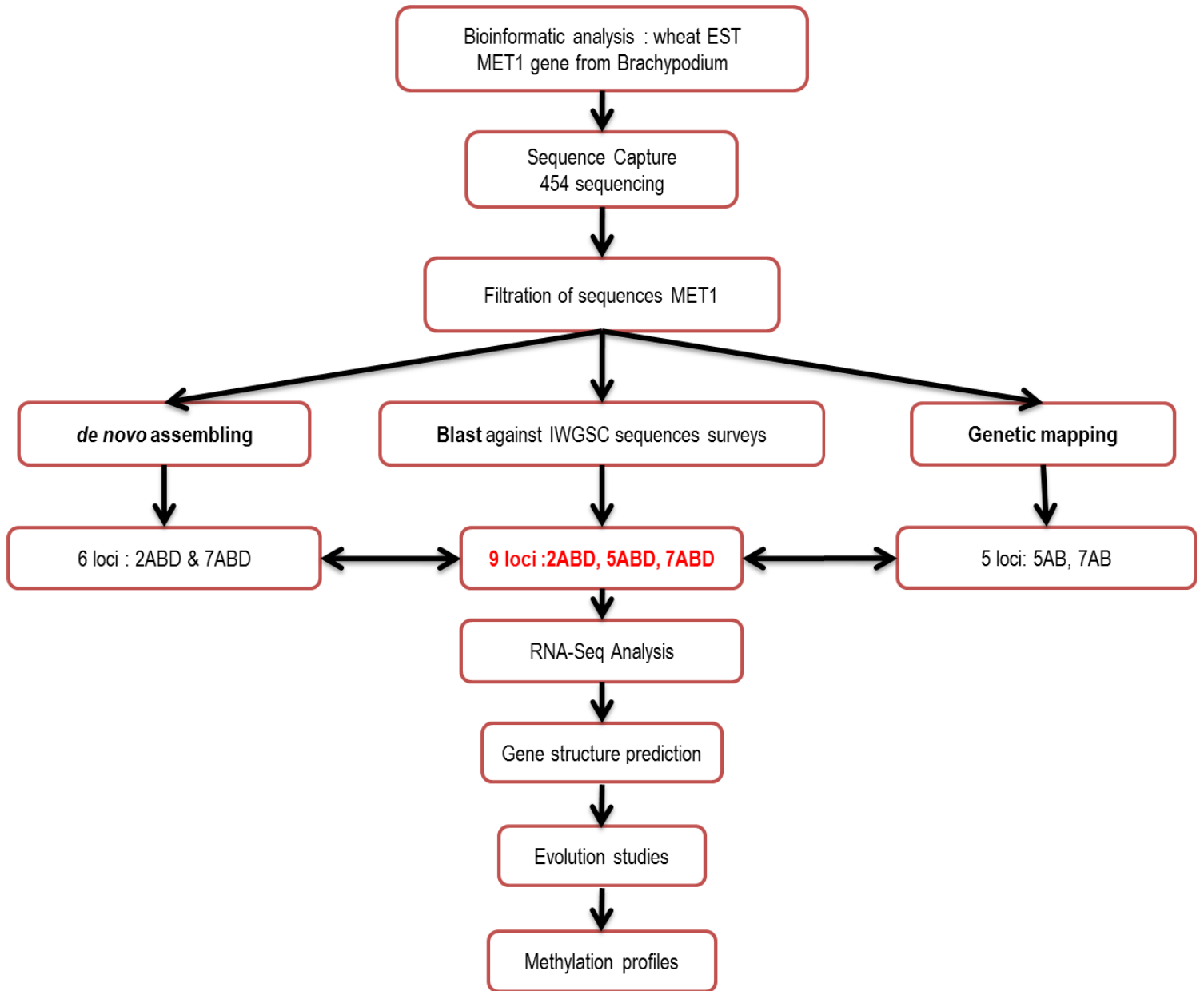


Figure 22: Schématisation de notre approche expérimentale

4. Article : Genome dynamics at Methyltransferase 1 genomic regions in hexaploid wheat reveals the involvement of CG DNA methylation in genome evolution

Soumission à BMC Genomics

Title page

Evolutionary history of Methyltransferase 1 genes in hexaploid wheat

THOMAS Mélanie^{1,2}

Email: melanie.thomas1@etudiant.univ-bpclermont.fr

PINGAULT Lise³

Email : lise.pingault@clermont.inra.fr

POULET Axel¹

Email : pouletaxel@gmail.com

DUARTE Jorge²

Email: jorge.duarte@biogemma.com

THROUDE Mickaël²

Email: mickael.throude@biogemma.com

FAURE Sébastien²

Email: sebastien.faure@biogemma.com

PICHON Jean Philippe²

Email: jeanphilippe.pichon@biogemma.com

PAUX Etienne³

Email : etienne.paux@clermont.inra.fr

PROBST Aline Valeska¹

Email: aline.probst@univ-bpclermont.fr

TATOUT Christophe^{1,4}

Email: christophe.tatout@univ-bpclermont.fr

¹UMR CNRS 6293 INSERM U 1103 Clermont Université, Genetics Reproduction and Development (GReD), 24 avenue des Landais, BP80026 63171 Aubière Cedex, France

² BIOGEMMA, route d'Ennezat, Centre de Recherche de Chappes, CS 90126, 63720 Chappes, France

³ UMR INRA 1095 Blaise Pascal University ,Genetics Diversity & Ecophysiology of Cereals (GDEC), Clermont-Ferrand – Theix, 5 chemin de Beaulieu, 63039 Clermont-Ferrand Cedex 2, France

⁴ corresponding author

Abstract

Background

Plant and animal methyltransferases are key enzymes involved in DNA methylation at cytosine residues, required for gene expression control and genome stability. Taking advantage of the new sequence surveys of the wheat genome recently released by the International Wheat Genome Sequencing Consortium, we identified and characterized *MET1* genes in the hexaploid wheat *Triticum aestivum* (*TaMET1*).

Results

Nine *TaMET1* genes were identified and mapped on homoeologous chromosome groups 2A/2B/2D, 5A/5B/5D and 7A/7B/7D. Synteny analysis and evolution rates suggest that the genome organization of *TaMET1* genes results from a whole genome duplication shared within the grass family, and a second gene duplication, which occurred specifically in the *Triticeae* tribe prior to the speciation of diploid wheat. Higher expression levels were observed for *TaMET1* homoeologous group 2 genes compared to group 5 and 7, indicating that group 2 homoeologous genes are predominant at the transcriptional level, while group 5 evolved into pseudogenes. We show the connection between low expression levels, elevated evolution rates and unexpected enrichment in CG-dinucleotides (CG-rich isochores) at putative promoter regions of homoeologous group 5 and 7, but not of group 2 *TaMET1* genes. Bisulfite sequencing reveals that these CG-rich isochores are highly methylated in a CG context, which is the expected target of *TaMET1*.

Conclusions

We retraced the evolutionary history of *MET1* genes in wheat, explaining the predominance of group 2 homoeologous genes and suggest CG-DNA methylation as one of the mechanisms involved in wheat genome dynamics.

Keywords

DNA methylation, evolution, genome dynamics, CG-rich isochores

Background

Triticum aestivum (hexaploid wheat or bread wheat) is one of the most important cultivated species in the world and it has been subject of intense research. Investigations of its genome structure led to the discovery of its highly dynamic nature during evolution. Using fossil records and phylogenetic studies, its evolution was traced from ancestral diploid species proposed to originate 50-77 million years ago (MYa) [1]. Indeed, bread wheat is a hexaploid species made of three homoeologous genomes called A, B and D which derived from different diploid species. These are proposed to be *Triticum urartu* ($2n = 2X = 14$ chromosomes, AA) and a diploid species related to *Aegilops speltoides* ($2n = 2x = 14$, BB) which gave rise some 0.5~0.6 MYa ago to *Triticum durum* ($2n = 4X = 28$ chromosomes, AABB). About 8,000 years ago, hybridization occurred between *Triticum durum* and *Aegilops tauschii* ($2n = 2X = 14$ chromosomes, DD) and yielded *Triticum aestivum* ($2n = 6X = 42$ chromosomes, AABBDD), the hexaploid wheat [2]. This means that every single gene is expected to be found in triplicate. The genome structure, organized in homoeologous genomes A, B and D, has to be maintained through cell division, a function which is ensured by the *Ph1* suppressor locus. The *Ph1* locus restricts homoeologous recombination and permits proper chromosome segregation in a hexaploid context through mitosis and meiosis [3].

Complementary approaches known as comparative genomics [4] at the genome-level (synteny) or the chromosome level (micro-synteny) were used to predict the genome structure of wheat in comparison to sequenced diploid species such as rice [5, 6], sorghum [7], maize [8], brachypodium [9] and more recently barley [10]. Recent syntenic studies proposed that the

ancestral genome of grass species was a diploid species organized in five chromosomes ($2n = 2X = 10$ chromosomes) [11]. From this initial chromosome organization, the ancestral diploid genome was duplicated through Whole-Genome Duplication (WGD) then fragmented giving rise to an intermediate ancestor with $2n=2X=24$ chromosomes [11]. This genomic structure has been well conserved in rice ($2n=2X=24$ chromosomes) while it evolved to $2n=2X=14$ through chromosome rearrangements in diploid wheat. Although WGD is expected to have had a large impact on wheat genome evolution it is not the only mode of genome rearrangement. Indeed, duplication of large chromosomal regions (segmental duplication), duplication at the gene level or tandem duplications have also occurred in the course of evolution [12]. Furthermore, it is now well established that wheat genome organization has been largely influenced by transposable element mobilization [13]. Most of the mechanisms described above increase genome size and lead to an elevated gene copy number. However, much less is known about reverse mechanisms, which reduce genome size to restore a diploid situation and reform single copy gene states. Indeed, early studies in *Saccharomyces cerevisiae* indicate that only 12% of the duplicate pairs remain after WGD suggesting that an extensive gene loss occurs after WGD [14]. In flowering plants, a fraction of single-copy genes were recently investigated and new hypotheses were suggested [15]: basically, after duplication, genes within one of the duplicated segments tend to be lost through small deletions while most genes are retained within the second segment, a mechanism known as fractionation bias [16]. Another difference occurring after duplication between two genomic segments is known as genome dominance during which one of the two segments shows higher expression levels than the other [16]. Data from maize and brassica further indicate that both gene fractionation, leading to extensive gene loss, and genome dominance are occurring simultaneously keeping the expression of the retained genes at elevated levels [16, 17]. Hexaploid wheat does not show an overall genome-wide transcriptional dominance of A, B or D subgenomes although some homoeologous genes can adopt a specific expression pattern [18]. All these recent outcomes reveal important genome dynamics, which affect genome size or organization and alter gene

expression. However, mechanisms implicated in these phenomena remain largely hypothetical, although epigenetic mechanisms have been suggested to provide means to induce asymmetric levels of expression between the two duplicated fragments prior to gene fractionation [16].

Although our knowledge about the hexaploid wheat genome structure is increasing, it remains challenging to decipher every step leading to its large genome size of about 16-17 Gb, which includes up to 80% of repeated sequences [13]. In polyploid genomes like cotton, rapeseed or wheat, several studies suggested the importance of epigenetic mechanisms in maintaining genome structure and chromatin stability as well as in regulating gene expression after hybridization and polyploidization [19, 20]. One of these epigenetic mechanisms is DNA methylation, which takes place at the carbon-5 cytosine residues in CG, CHG and CHH (where H = A, T or C) contexts [21]. Loss of DNA methylation causes reactivation of silenced transposable elements [22] and the expression of certain genes, such as *FWA*, a gene involved in flowering [23, 24]. DNA methylation is also known to affect crossover rate and meiotic recombination [25].

We wanted to reconstruct the evolutionary history of the hexaploid wheat species *Triticum aestivum* using the example of *MET1*, a gene encoding METHYLTRANSFERASE 1 (MET1), responsible for DNA methylation maintenance in the CG context. *MET1* is a gene of particular importance for genome maintenance in many organisms, which we hypothesize, will be a crucial component of epigenetic mechanisms controlling transposable elements that in wheat make up to 80% of the genome. To date *MET1* gene function have been described in several plant species including Arabidopsis [26], maize [27], rice [28] and brassica [29] but not in wheat. We identified *MET1* genes in hexaploid wheat (*TaMET1*). Nine copies of *TaMET1* are organized in three paralogous groups at chromosome 2, 5 and 7 suggesting that the genomic regions including *MET1* genes were subjected to two duplication events prior to the emergence of hexaploid wheat.

Considering *TaMET1* genomic regions specifically, we confirmed that the first gene duplication was part of a WGD common to all grass species and that the second duplication occurred through gene duplication specific to the *Triticeae* tribe. Expression profiles of the different *MET1* gene

copies, estimation of their evolution rates, CG enrichment and methylation profiles highlight the predominance of group 2 homoeologous genes at the transcript level. Our results exemplify the high dynamics of genome evolution in the course of the evolutionary history of bread wheat and suggest the involvement of epigenetic mechanisms in these processes.

Results

Hexaploid wheat contains nine *TaMET1* loci

In order to determine the number and complete sequence of *TaMET1* genes, we chose a genomic strategy based on a combination of sequence capture and *in silico* mining of available wheat genome sequences. In order to define probes on the sequence capture microarray, *TaMET1*-expressed tags (ESTs) were identified in wheat databases. Eight ESTs were retrieved from public and private libraries. The alignment of these ESTs with the rice and brachypodium *MET1* genes showed that these ESTs mapped to the 3' end of *TaMET1* genes. Two *TaMET1* ESTs as well as two brachypodium *MET1* genomic fragments were selected and used to design probes for sequence capture (see Methods). Two successive runs of sequence capture yielded 8,184 reads specific to *TaMET1* genes. Genomic fragments were then assembled *de novo* using gsAssembler in six large contigs corresponding to six putative *TaMET1* genes. However some reads remained impossible to assemble and could not be included within the six large contigs suggesting the possible existence of additional copies of *TaMET1*. In parallel, the draft genome assembly of the wheat genome recently released by Brenchley and collaborators [30] was mined for *TaMET1* genes. However, no full-length sequences corresponding to *TaMET1* genes were present in the dataset. Taking advantage of the recent release of sequence surveys from the International Wheat Genome Sequencing Consortium (IWGSC) (<http://www.wheatgenome.org/>) that were produced from sorted chromosome arms [18], BLASTn analyses against each chromosome arm were

performed using rice and brachypodium *MET1* genes. Eventually nine *MET1* copies were identified and assigned to chromosomes 2A/2B/2D, 5A/5B/5D and 7A/7B/7D. For simplicity, homoeologous chromosomes A, B and D will be collectively referred to as a “homoeologous group” hereafter. Intron and exon junctions were defined for the nine *TaMET1* genes according to rice and brachypodium *MET1* genes and subsequently validated by RNA-Seq analysis (see below). Protein domains were then predicted using the Pfam database. Three major protein domains were identified that include DNMT1-RFD (Cytosine specific DNA methyltransferase replication foci domain), BAH (Bromo-Adjacent Homology) and the DNA methyltransferase (C5 cytosine specific DNA methylase) domain (figure 1). Comparison with *MET1* genes from rice orthologs showed an overall conservation of the *TaMET1* genes (figure 1). Coding sequence analyses revealed that the *TaMET1* genes of chromosome 5A and 5D display deletions and premature stop codons (figure 1) and if expressed produce truncated proteins missing the DNA methyltransferase domain. *TaMET-5A1* and *-5D1* may be considered as pseudogenes, while all the remaining genes are expected to be functional.

***TaMET1* loci originated from two successive duplication events**

In a first attempt to understand the genome rearrangements, which have led to the nine *TaMET1* genes present in the *T. aestivum* genome, we retraced the phylogenetic history of *TaMET1* genes using *MET1* orthologs from monocotyledonous and dicotyledonous species. Two distinct copies of *MET1* (*i.e* two distinct paralogs) are usually found in monocots such as rice, sorghum and brachypodium species. Phylogenetic analysis suggests that homoeologous *TaMET1* genes from group 2 are orthologous to *OsMET1b* on chromosome 7 while homoeologous *TaMET1* genes from group 5 and 7 are orthologous to *OsMET1a* on chromosome 3 (figure 2A). Hereafter, these two phylogenetic groups are called MET-1a and MET-1b lineages in respect to the *MET1* genes from rice. The phylogenetic tree suggests that a first duplication event occurred early during monocot speciation resulting in the MET-1a and the MET-1b lineages (figure 2A). Since these two copies

of *MET1* are common to all grass species, the first *TaMET1* duplication is likely to be a consequence of the WGD that took place in all grasses and occurred about 56-73 MYa. The second duplication is shared only within the *Triticea* tribe (barley and wheat in our phylogenetic tree). Since wheat diverged from brachypodium 32-39 MYa and from barley 10-13 MYa [1, 9], this second duplication probably occurred between 32 and 13 MYa. In order to understand if this duplication was the result of segmental or single gene duplication, syntenic relationships between regions surrounding the *TaMET1* genes from chromosome 5 and 7 and their orthologous loci in rice and brachypodium were investigated. For chromosome 5A, 5B and 5D, up to 80% of the genes were conserved, whereas only 10-15% were for group 7, suggesting that a single gene duplication occurred (figure 2B). This hypothesis is consistent with the evolutionary model of grass genomes [11, 31].

In order to date the duplication event leading to group 5 paralogs, BLASTn analyses were conducted between hexaploid wheat (*Triticum aestivum*), diploid wheat species (*Triticum urartu*, *Aegilops tauschii*) and barley (*Hordeum vulgare*). *Triticum urartu* (genome A ancestor) shares the same deletion with *TaMET-5A1* while *Aegilops tauschii* (genome D ancestor) and *TaMET-5D1* do not (figure 2C). It can then be suggested that 5A was already in the process of pseudogenization before polyploidization while 5D pseudogenization occurred in the course of, or after, polyploidization. Consistent with this hypothesis, 5A displays a more pronounced gene structure alteration than 5D (large deletion and numerous stop codons; see also figure 1).

***TaMET1* genes display distinctive evolution rates**

We then investigated the putative functional differences between the nine *TaMET1* genes by evaluation of the evolution rate, which is a good indicator for the biological function of a given gene [32]. We chose the codon substitution model to estimate the rate of synonymous (dS) and non-synonymous (dN) substitutions and computed the dN/dS ratio as evolution rate (ω) [33]. In this model, for genes with a significant

biological function undergoing purifying selection non-synonymous mutations are expected to be kept at a low level whereas synonymous mutations accumulate randomly according to the neutral theory of evolution [33].

As a first approach, pair-wise divergences were investigated between *TaMET1* genes and *MET1* genes from fully sequenced monocot species (*i.e.* divergence between orthologous pairs). Mean values for A, B and D homoeologs were then calculated per homoeologous group of chromosomes (group 2, 5 and 7) and are displayed in figure 3A. Consistent with the neutral theory of evolution, dS rates were not significantly different between the three homoeologous groups. However significant differences were observed for dN and ω indicating a lower rate of evolution for homoeologous group 2 which belongs to the *MET-1b* lineage. Homoeologous group 7 is evolving at an intermediate evolution rate compared to group 2 and group 5 but does not display any deleterious mutations within the coding sequences (see also figure 1). As expected for pseudogenes, higher dN and ω values were found for *TaMET1* at homoeologous group 5.

Secondly, various hypotheses concerning evolution rates were then tested and Likelihood Rate Tests (LRT) were computed. Eleven hypotheses were evaluated to test an increased evolution rate at various branch points in the phylogenetic tree (additional file 1). Evolution rates are summarized in figure 3B. The results support the existence of three evolution rates (indicated as ω_0 , ω_1 and ω_2 in figure 3B) consistent with the two duplication events and the pair-wise analysis performed previously (figure 3A). After gene duplication, long-term changes were observed in our phylogenetic tree. ω_0 , ω_1 and ω_2 evolution rates indicate that negative selection occurs in the *MET-1b* lineage, which has the smallest evolution rate ($\omega_0 = 0.12193$) suggesting its functional role in monocots. Following the first duplication event, a two fold increase in evolution rate ($\omega_1 = 0.21988$) is observed in the *Met-1a* lineage except for barley chromosome 5 and wheat homoeologous group 5 for which a fourfold increase ($\omega_2 = 0.48706$) is observed.

Altogether, evolution rate analyses indicate that *TaMET1* homoeologous genes of group 2 are submitted to stronger purifying selection and are evolving at a slower rate suggesting their predominant role in DNA methylation maintenance in hexaploid wheat. Following the second duplication event, asymmetric acceleration of the evolutionary rate is observed between homoeologous group 5 and 7 leading eventually to the formation of pseudogenes within group 5 that accumulated deleterious mutations within their coding sequences.

Expression of *TaMET1* genes is mainly driven by homoeologous group 2

The above analysis of evolution rates suggests that homoeologous genes from group 2 are under purifying selection. As it is well documented that expression patterns influence non-synonymous substitution [34], expression levels and profiles of the nine *TaMET1* genes were investigated using RNAseq data from five different organs at three developmental stages each. Expression levels of *TaMET1* homoeologous genes from group 2 were found to be 10 to 40 times higher than the *TaMET-5B1* and *TaMET* group 7 ones. For *TaMET-5A1* and *5D1*, no significant expression was detected in any of the 15 conditions (figure 4A). Homoeologous group 2 were expressed in most tissues at nearly all developmental stages, named according to the Zadoks (Z) scale [35], but with highest expression levels at Z30 in the stem and Z32 in the spike. MET1 expression levels in other species peak in proliferating cells such as in meristems and in reproductive organs [27, 28, 36]. In wheat we observed *TaMET1* expression at early developmental stages especially during early stem extension (Z30-Z32) when wheat is switching from the vegetative to reproductive phase. At that stage the spike tissue is actively proliferating requiring active replication during which DNA methylation maintenance should occur. Similarly, homoeologous group 7 were found to be expressed in almost all conditions but at a very low level compared to group 2 genes. A similar situation was observed in rice where *OsMET1a* is 10-12 times less expressed than *OsMET1b* [37]. For homoeologous group 5, only 5B is expressed at low level in grain (figure 4A). RNAseq-based expression profiles were subsequently confirmed by RT-PCR. Various primer pairs were designed at the 3'UTR. Semi-quantitative and quantitative analyses confirmed the expression of *TaMET1* from homoeologous group 2 (figure 4B and 4C) but transcripts were hardly or not detectable for group 5 and 7 (figure 4B). Expression levels for 2A, 2B and 2D homoeologs were investigated by RT-qPCR but did not show strong differences, although 2B was found to be slightly more strongly expressed (figure 4C). Thus *TaMET-2A1*, *-2B1* and *2D1* are expressed in an additive mode.

Recent analyses at the whole genome level indicated that housekeeping genes in wheat are enriched at pericentric positions while genes expressed with tissue-specific patterns and pseudogenes are usually found at more sub-telomeric positions [38]. To check whether there is a correlation between the observed gene expression differences and the physical position of *TaMET1* copies on the chromosomes, we genetically mapped *TaMET1* loci using 57 SNPs identified in the course of our sequence capture experiments (see Methods). Out of the 57 SNPs, 18 produced high quality results that led to the genetic mapping of five out of nine *TaMET1* genes, namely *TaMET-2B1*, *5A1*, *5B1*, *7A1* and *7B1*. As positions of homoeologous copies were consistent for groups 5 and 7, we extrapolated the position of all *TaMET1* genes from these five copies. Homoeologous group 2 were found to be located in the pericentromeric regions of the short arm of chromosomes 2 whereas group 5 and 7 were mapped in subtelomeric positions of the long arms of chromosomes 5 and 7 respectively (figure 4D and additional file 2).

Thus MET1 expression is mainly driven by homoeologous group 2 indicating specific mechanisms to keep a predominant expression of homoeologous group 2 over groups 5 and 7. This observation resembles a phenomenon observed after *MET1* gene duplication in Arabidopsis where *MET1* transcripts accumulate to 10,000 fold higher levels than those of the duplicated *MET1a* and *b*, while *MET1c* is considered to be a pseudogene [39]. Expression of a specific member of a given gene family is referred to as predominance [40] or transcriptional dominance [16] and in our case occurs for *TaMET1* genes at homoeologous group 2. The pericentric position of group 2 genes is consistent with the conclusions drawn from a recent large scale analysis indicating that genes expressed in most tissues are located in more proximal position than those displaying tissue-specific expression patterns [38]. Thus expression studies reinforce the idea that *MET1* homoeolous group 2 genes might provide methyltransferase activity.

CG-rich isochores appear at *TaMET1* promoters and exhibit high DNA methylation

While low, or absent expression of specific *TaMET1* genes might be explained by several factors including genetic mutations or insertion of transposable elements, epigenetic marks at promoter regions are good candidates to explain differences in gene expression [23, 24, 41, 42]. Among

these, cytosine methylation that occurs in CG sequence contexts has been shown to modulate gene expression in plants [23, 24, 41, 42]. To investigate the potential role of DNA methylation in the regulation of the *MET1* genes, *MET1* coding sequences as well as putative promoters were scanned for potential methylation in CG, CHG and CHH sequence contexts.

The putative promoters of the nine genes were defined as ~2-3kb regions upstream of the coding sequence depending upon the availability of the genomic sequences (figure 1). Comparisons between upstream and coding sequences for potential methylation sites in CHG and CHH contexts revealed similar amounts of CHG and CHH sites for all nine genes (data not shown). In contrast, cytosines in the CG context were enriched at potential promoter regions of homoeologous group 5 (4.4 fold) and group 7 (5.5 fold) compared to group 2 putative promoters regions (figure 5A and 5B). This result was unexpected because CG-rich regions (also known as CG-rich isochores), although already described in Arabidopsis genes, were shown to be mainly located in introns [43]. As CG-rich isochores at *TaMET1* upstream regions could be the result of new insertions of CG-rich DNA sequences, we looked for such events. Indeed, two DNA insertions of 786 and 122bp overlapping with CG-rich isochores were observed for the *TaMET-7D1* upstream region (additional file 3). Both insertions were already present within the ancestral D genome (*Aegilops tauschii*) suggesting their integration prior to polyploidization (additional file 3). BLASTn analysis against the TREP database indicated a short but significant homology with a *stowaway* Miniature Inverted Repeat (MITE) for the 122bp insertion while no significant homology was detected for the larger 786bp insertion. BLASTn against TREP performed with the five remaining upstream regions (7A, 7B, 5A, 5B and 5D) failed to detect any transposable elements as shown in figure 1. Instead of a new large DNA insertion enriched in CG observed at 7D, the CG-rich isochores are more dispersed along the 5A, 5B, 5D, 7A and 7B putative promoter regions (figure 5A). This may argue in favor of a progressive CG accumulation in the course of evolution.

To determine whether these regions enriched in cytosine residues on homoeologous group 5 and 7 are indeed methylated, we performed bisulfite sequencing. We designed bisulfite primers in a way

to simultaneously amplify all three homoeologous copies that we subsequently discriminated upon sequencing. Consistent with our expression studies, putative promoter regions from homoeologous groups 5 and 7 display DNA methylation in CG sequence contexts. Homoeologous group 5 also displays significant CHG methylation (figure 5C and additional file 4 and 5). Among all the analyzed putative promoter regions, the highest DNA methylation levels overlap with the 786bp insertion specific to 7D (figure 5C and additional file 5).

Taken together, our results suggest that the presence of CG-rich isochores in the putative promoters of group 5 and 7 *TaMET1* homoeologous genes may be due to a progressive and dispersed CG-enrichment as well as to an insertion-mediated CG-enrichment, at least for the 7D copy. In addition, the high methylation levels observed in the promoter regions of the two low-expressed homoeologous groups may suggest the existence of an autoregulatory loop controlling *MET1* gene expression.

Discussion

Bread wheat is a plant species with a large genome of about 17 Gb containing up to 80% of repetitive sequences. Much attention has been focused recently to understand how this genome, highly enriched in repetitive sequences, controls its transposable element fraction, which will otherwise lead to genome instability. One such mechanism is likely to involve DNA methylation in the CG context, which is maintained by MET1. It is therefore of importance to understand how MET1 expression is regulated in an organism with a complex hexaploid genome. In the course of our work, we observed that *TaMET1* genes contain a record of many evolutionary events, which have occurred prior and after the emergence of bread wheat.

We identified nine copies of *TaMET1* organized in three homoeologous groups at chromosomes 2, 5 and 7. At the chromosomal level, segments bearing *TaMET1* originated from two duplication events. Phylogenetic- and micro-synteny confirmed that chromosome 2 and 5 paralogs originated

from a WGD about 50-70 MYa in the ancestor of grass species. Then the chromosome 7 paralog emerged from a more recent gene duplication about 13-32 MYa in the *Triticea* tribe. Our analysis of the evolution rate revealed functional differences between the nine *TaMET1* genes. The MET-1b lineage (homoeologous group 2) was shown to display a lower evolution rate than the MET-1a lineage (homoeologous group 5 and 7). Lower evolution rate is observed for genes with biological function and this is best explained by purifying selection, which counter selects deleterious mutations [34, 44]. Functional significance of homoeologous group 2 genes was reinforced by our observations of expression levels and DNA methylation. Low evolution rate in the MET-1b lineage matches with a predominant expression of homoeologous group 2 over group 5 and 7. Predominant expression of one member of the *MET1* gene family was already observed in other species such as *Arabidopsis* [39] and rice [37] suggesting that MET1 expression level and pattern needs to be carefully controlled. Interestingly, we mapped *TaMET1* homoeologous group 2 to peri-centric (proximal) position while group 5 and 7 were located at more sub-telomeric (distal) regions. Recent large scale analyses in wheat suggested that distal regions are more dynamic, displaying higher level of recombination and accumulate more pseudogenes and gene duplications than proximal peri-centric regions [38]. Furthermore, genes at distal position display more tissue specific expression than those at more proximal position. It is then tempting to hypothesise that a distal chromosome position may have a direct influence on expression leading as a consequence to the predominance of the more proximal genes as observed in our case for homoeologous group 2. Homoeologous group 2 did not show any differences in gene expression among the three homoeologs. Consistent with the whole genome analyses was the fact that genome-wide transcriptional dominance of an individual subgenome (A, B or D) was not observed [18]. Besides its position along the chromosome, our data indicated that DNA methylation observed in the promoter region of homoeologous group 5 and 7 may have contributed to their transcriptional repression and may have favored an increased evolution rate at *TaMET-5A1* and *5D1* leading to the accumulation of deleterious mutations, a process known as pseudogenization [45, 46].

Interestingly, distinctions can be made between group 5 homoeologs: 5A already accumulated large deletions and numerous stop codons before polyploidization, while stop codons occurred at 5D after polyploidization but are absent at 5B which however displays an elevated level of non-synonymous mutations and is expressed only in grains. Pseudogenes are usually rapidly eliminated and the fact that *TaMET-5A1* and *5D1* pseudogenes are maintained suggests that pseudogenization may not be fully achieved or that these genes contribute in a significant but yet unknown manner to *TaMET1* expression.

Our data support a functional role of DNA methylation in the initiation or the maintenance of gene silencing at specific *TaMET1* genes. Considering that the chromosome 2 paralog is the ancestral locus and shows low occurrence of potential CG methylation sites, the observed CG-rich isochores at chromosome 5 and 7 paralogs associated with DNA methylation imply CG-enrichment at these putative promoter regions. CG-enrichment was unexpected as usually CG dinucleotides are under-represented due to 5-methylcytosine deamination [43]. At the moment we can only speculate about their possible origin. First, GC-rich and GC-poor isochores are known to occur in animals and several hypotheses have been proposed to explain their emergence [47]. Among them the GC-biased gene conversion (gBGC) has been proposed as one of the main driving forces in the evolution of nucleotide composition. gBGC favors GC over AT bases in alleles during mismatch repair following heteroduplex formation in the course of meiosis. gBGC results from Base Excision Repair (BER) and involves a DNA glycosylase that specifically removes thymine in DNA heteroduplexes. Secondly, animal genomes display unmethylated CG-rich elements known as CpG islands (CGIs). CGIs are defined as DNA sequences of a few hundred base pairs, with high CG occurrence, high G+C frequency and are involved in the regulation of gene expression [48]. CGIs have been divided into start and non-start CGIs. Non-start CGI are the most abundant and best explained by insertion of repeated sequences such as transposable elements (in the human genome 79% are due to *Alus*) while start CGIs located at the transcription start sites are only poorly explained by transposable element insertion (in the human

genome 5,6% are due to *Alus*) [49]. Interestingly, Suzuki et al., 2011 [50] also proposed gBGC as one of the possible mechanisms to explain the emergence of start CGIs. Recently, it was suggested that gBGC occurs in plants [51]. gBGC can be considered as one of the possible mechanisms explaining the emergence of CG-rich isochores at *TaMET1* putative promoter regions. Indeed, it may be an attractive mechanism to explain the progressive CG enrichment we observed at *TaMET1* upstream regions especially at homoeologous group 5 and 7 located at distal chromosome positions where higher recombination rates have been reported [38, 52]. Furthermore, the *MET-7D1* copy would have also undergone insertion of CG-rich DNA fragments in a mechanism very reminiscent to the one observed for non-start CGIs, arguing for shared evolutionary mechanisms between animal and plants toward the emergence of CG-rich isochores. Once CG-rich isochores appeared, they can be methylated in order to silence gene expression. Although CGIs were not described in plant promoters, “dense CG methylation clusters” have been reported and are proposed to silence cryptic promoters within the coding sequence [43]. Silencing of these cryptic promoters is established first through the RNA-directed DNA Methylation (RdDM) pathway and results in methylation at cytosine residues at CG, CHG and CHH sequence contexts. Once methylation is set up, only methylation in the CG context, which does not rely on siRNA production, can be maintained in the course of evolution leading to high methylation only in CG sequence contexts [43]. If such a mechanism occurred within the putative promoter region of *TaMET1* genes, it can explain how homoeologous group 7 became progressively repressed. Given the correlation between DNA methylation in promoter regions and gene silencing [23, 24], we suggest that DNA methylation may be part of a possible auto-regulatory mechanism among *TaMET1* genes. In this model, MET1 mainly encoded by homoeologous group 2 regulates group 7 gene expression through CG DNA methylation maintenance. CG methylation at homoeologous group 7 may be alleviated in specific organs, developmental stages or upon changing environmental conditions. However possible roles for the homoeologous group 7 (MET1-a lineage) is challenged by recent data collected in rice indicating that the main MET1 function is

ensured by *Met1b* and not *Met1a*. Indeed, RNAi against *Met1a* does not significantly affect plant development while a *met1b* null mutant is lethal [28, 53].

Conclusions

From our data, we propose a chronology (figure 6) of the genomic events observed at *TaMET1* genes, which include WGD, gene duplication, expression predominance of homoeologous group 2, CG-rich isochores emergence, DNA methylation and pseudogenization. The unexpectedly rich evolution history observed at *TaMET1* makes these loci a very attractive model to study further gene evolutionary mechanisms occurring in hexaploid wheat. Increased copy number finally leads to *TaMET1* silencing at homoeologous group 5 and 7 (the MET-1a lineage), keeping genes of group 2 (the MET-1b lineage) in an active state. We hypothesize that CG methylation was used as a mean to control gene expression in the MET-1a lineage favoring low expression at homoeologous group 7 and pseudogenization at group 5. For the latter the different evolutionary stages are still observed between homoeologs. CG methylation might be required to limit homoeologous group 7 transcription using CG-rich isochores, which are the target of CG DNA methylation. At that point, we can only speculate about the possible involvement of methylation in limiting homoeologous group 7 expression in tissues or developmental stages where group 2 is expressed, maybe leading to tissue-specific expression patterns of group 7 genes and their subfunctionalization.

Taken together, our data suggest that DNA methylation at *TaMET1* loci can act as an epigenetic determinant required to drive genome evolution.

Methods

Analysis of IWGSC Sequence surveys

Access to the IWGSC physical map sequences from hexaploid wheat *cv Chinese Spring* has been established by the URGI (<http://urgi.versailles.inra.fr/>) on the behalf of IWGSC. BLASTn analyses were performed at the URGI database. Identification of transposable elements in selected IWGSC contigs was performed by BLASTn analyses against the TREP database at <http://wheat.pw.usda.gov/ITMI/Repeats/>. CG, CHG and CHH profiles (where H is A, C or T) were detected using an in house Perl script available upon request. Gene structures were predicted by cDNA alignment against genomic sequences using *SIM4* (<http://pbil.univ-lyon1.fr/>); protein domains are according to the Pfam database (<http://pfam.sanger.ac.uk/>). Gene structures were subsequently designed by *FancyGene* (<http://bio.ieo.eu/fancygene/>).

Sequence Capture

Sequence capture experiments [54, 55] were designed to isolate and sequence DNA segments using probes synthesized on a microarray. Two runs of sequence capture were performed according to the NimbleGen Arrays User's Guide followed by 454 Optimized Sequence Capture method. Briefly, among the whole array design, *MET1* specific probes ranging from 60 to 90 nucleotides were designed at a density of $\sim 1.10^6$ probes / Mb of sequence by Roche-Nimblegen from one *Triticum aestivum* public EST (accession number TA8302), one private EST (GPIC:S:720428) and from the two coding sequences from *Brachypodium distachyon* (Accession numbers bradi1g05380 and bradi1g55290). The absence of repeated sequences was verified using repeatMasker and TREP release 10 database. Genomic DNA from wheat elite lines *cv Brigadier*, *Alcedo*, *Renan* and *Recital* were used to build up four distinct genomic DNA libraries by nebulization with an average of fragment sizes of ~ 600 bp. Libraries were then hybridized onto capture arrays, captured DNA was eluted and amplified prior to 454 sequencing on a GS FLX Titanium platform according to the manufacturer. Overall, sequence captures yielded 8,184 reads specific to *MET1* and *de novo* assembly was subsequently performed using *gsAssembler* (Roche) with specific parameters set at 98% similarity and 20bp overlap. Sequences were verified in the

course of the project by BLASTn analysis against the IWGSC surveys and by PCR amplification on diploid and hexaploid species.

Genetic Mapping

Two mapping populations were used: recombinant Inbred Lines derived from a cross between *Triticum aestivum* cv *Renan* and *Recital* [56] and a doubled haploid population derived from a cross between *Triticum aestivum* cv *Brigadier* and *Alcedo* (Biogemma personal communication). DNA from all four elite lines was used in the sequence capture experiments and reads were grouped according to the four DNA origins. As a whole, 57 putative Single Nucleotide Polymorphisms (SNPs) were identified and genotyping was subsequently performed on genomic DNA from the two mapping populations using KASPar (KBioscience) fluorescent competitive allele-specific amplification. Primers were designed with *Primer picker* (KBioscience) and PCR amplifications were performed on a hydrocycler (LGC genomics), for 41 to 50 cycles at 57°C and then run onto a Genotyper (Applied Biosystem). The list of primers used to perform the genetic mapping is provided in additional file 6. SNP mapping was performed on the two genetic maps using an in-house bioinformatic pipeline available at Biogemma. Genetic positions are given according to the *Renan* x *Recital* recombination map [56]. Physical positions are according to the names of the IWGSC contigs obtained by BLASTn analysis against the virtual map and are included within the virtual map designed by synteny.

Phylogenetic reconstruction and substitution rate calculation

TaMET1 coding sequences were used for phylogenic reconstruction and substitution rate calculation.

Selected sequences were first aligned with *MUSCLE* multiple sequence alignment [57] and then refined using *Gblocks* [58]. Maximum likelihood analysis was performed with *PhyML* using default parameters with 1,000 bootstraps [59]. Phylogenetic trees were drawn using *ITOL* [60].

Substitution rate studies were performed as follows: first, a new phylogenetic tree was built with the same species except that here, dicot species were not considered and the tree was based on protein sequences instead of cDNA. For *TaMET1*, genomic sequences were used to predict exonic sequences using *FGENESH* [61] and subsequently assembled into a predicted cDNA. Predicted cDNAs were validated in the course of this study by RNA-Seq data. cDNAs were translated using *Transeq* and *Sixpack* from the *EMBOSS* package [62]. The phylogenetic tree was then built from predicted proteins as described above. ω (the ratio of nonsynonymous/synonymous substitution rates) was determined using *Codeml* from the *PaML* package [33]. A likelihood ratio test (LRT) was used to compare various hypothesis models in which ω values are expected to differ among branches, in comparison to a null hypothesis in which all the branches have similar ω . LRT values were then compared to a chi-squared distribution with degrees of freedom equal for a given tree to the number of values of $\omega - 1$, as described in Yang 2007. Matrix and trees are available at <http://purl.org/phylo/treebase/phylows/study/TB2:S16421>.

Micro-synteny analyses

Starting from Murat *et al* 2010, chromosomal segments including *MET1* loci were selected from rice and brachypodium. To be able to compare our results with those of Murat *et al* 2010 [30], the same fragment boundaries were retained but in our case, all the coding sequences of a given genomic fragment have been considered. Briefly, rice chromosome 3 from [Phytozome: LOC_Os03g58040.1 to LOC_Os03g58920.1] (covering 510.70kb of genomic DNA and including 80 genes) and brachypodium chromosome 1 from [Phytozome: Bradi1g05680 to Bradi1g04980] (covering 531.9kb of genomic DNA and including 72 genes) chromosomal segments are syntenic to wheat chromosome 5 and 7 (additional file 7). Gene sequences from model species were then used to perform BLASTn analysis against the IWGSC sequence surveys as described in Salse *et al.* 2008 using 70% CIP (Cumulative Identity Percentage) but only 30% CALP (Cumulative Alignment Length Percentage). The CALP parameter was kept at a low value in order to detect all

the micro-syntenic relationships. Percentage of syntenic conservation was then computed as $100 \times$ the number of conserved genes in wheat / number of genes in the syntenic region from brachypodium or from rice.

RNASeq

RNA-seq non-oriented libraries were constructed in two replicates from total RNAs of hexaploid wheat *cv Chinese Spring*. RNAs were prepared with the TruSeq kit (Illumina) for 15 biological samples including 5 organs (root, leaves, stem, spike, grain) and 3 developmental stages (beginning, middle, and end of development) as described in [63](additional file 8). For oriented libraries, samples were pooled by organs, rRNAs were removed from total RNAs with the riboZero kit (Ambion) and RNA-seq libraries were constructed with the ScriptSeq kit (Epicentre). All the libraries were sequenced using a HiSeq200 (Illumina) with reads of 100 bp sequenced in both directions. RNA-Seq data have been deposited in the Sequence Read Archive (SRA) at <http://www.ncbi.nlm.nih.gov/sra/> under accession number ERP004714. Reads from RNA-seq libraries were mapped using *Tophat2 v2.0.8* [64] and *Bowtie2* [65] onto the *MET1* genomic sequences with neither mismatches nor splice-mismatches allowed. Transcript reconstruction and expression levels were analyzed with *Cufflinks v2.0.2* [66] without annotation. Because sequencing was bidirectional, which is to say that two reads correspond to the same cDNA molecule, expression data results of transcription levels are expressed in Fragments per Kilobase of Exon Model (FPKM) per million mapped reads [67].

RNA analyses

Wheat plantlets of *cv Chinese Spring* were grown in a greenhouse and collected at Z61-65 stage according to Zadoks scale [35]. Tissues were frozen in liquid nitrogen and ground to a fine powder. Total RNAs were extracted from 250mg of plant material using an RNA extraction method adapted from [68]. RNA was subsequently treated with 100 units of DNase I (Invitrogen)

in the presence of 20U RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen). Quantity of extracted RNA was estimated using a Nanodrop (Thermo Scientific) and RNA quality was checked by migration on a 2% agarose gel containing MOPS 2% and 0,05% formaldehyde. Reverse Transcription was performed from 2µg of total RNA using an oligo(dT) 15 Primer and M-MLV Reverse Transcriptase (Promega) in presence of Recombinant RNasin Ribonuclease Inhibitor (Promega) according to the supplier's recommendation. Homoeologous specific primers were designed manually and validated with *Oligo Analyzer* (Gene Link) to avoid secondary structure formation. Sequences of selected primer pairs can be found in additional file 6. Semi-quantitative analyses were performed using primer pairs with similar efficiencies and on the same cDNA sample by comparing the *TaMET1* PCR product to Ta4045 and Ta54227 as reference genes (primer pairs as as in [69]). Quantitative analysis was performed on a LightCycler® 480 System using LightCycler® 480 SYBR Green I Master reagent (Roche) according to the supplier's recommendation. Primer pair efficiencies were calculated through serial dilutions from 1/3 to 1/81 for each RNA sample and only primer pairs with a PCR efficiency between 80 and 110% were selected. As in semi-quantitative analyses, Ta4045 and Ta54227 were used as reference genes.

Bisulfite sequencing

1g of plant material was collected from stem and leaves at the Z30 stage and DNA extracted using the DNeasy plant maxi kit (Qiagen). 200-500ng of DNA was subjected to bisulfite (BS) treatment using the EZ DNA Methylation-Gold™ Kit (Zymo Research). BS-treated DNA was PCR-amplified using specific primers (additional file 6) and cloned in pGEMT vectors (Promega) prior to sequencing. 10-20 clones were analyzed for each genomic region using Kismeth software [70]. Two PCR fragments from the VERNALIZATION1 (*VRN1*) gene previously studied by bisulfite experiments [71] were used as internal controls. Incomplete conversion was excluded by analyzing the 0.0 k fragment from *VRN-A1*, which is devoid of CG methylation, while optimal

bisulfite treatment were assessed by analysis of the 9.2 k fragment, a highly CG methylated region from *VRN-A1*. Examples of results are given in Additional file 9.

Availability of supporting data

The following additional data is available with the online version of this paper. Additional data file 1 is a table listing the results of the Likelihood ratio tests. Additional data file 2 is a table listing the genetic positions of *TaMET1* loci. Additional data file 3 is a sequence alignment of the promoter region of *TaMET1* from homoeologous group 7 with close species. Additional data file 4 and 5 are detailed bisulfite analyses performed at *TaMET1* from homoeologous group 5 and 7 respectively. Additional data file 6 is a table listing the primers used in this study. Additional data file 7 is a table describing micro-synteny data between wheat, rice and brachypodium. Additional data file 8 is a table listing the RNA-Seq samples used in this study. Additional data file 9 is an example of control experiment in bisulfite sequencing analysis.

List of abbreviations

BAH: Bromo-Adjacent Homology; BER: Base Excision Repair; CALP: Cumulative Alignment Length Percentage; CIP: Cumulative Identity Percentage; CGIs: CpG islands; DNMT1-RFD: Cytosine specific DNA methyltransferase replication foci domain; LRT: Likelihood Rate Tests; ω : Evolution rate; ESTs: Expressed Sequence Tags; FPKM: Fragments per Kilobase of Exon Model; gBGC: GC-biased gene conversion; IWGSC: International Wheat Genome Sequencing Consortium; MET1: METHYLTRANSFERASE1; MYa: Million Years ago; MITE: Miniature Inverted Repeat; dN: rate of non-synonymous substitution; dS: rate of synonymous substitution;

RdDM: RNA-directed DNA Methylation; SNP: Single Nucleotide Polymorphism; WGD: Whole-Genome Duplication; Z: Zadoks scale.

Competing interests

The authors declare no conflict of interest.

Authors' contributions

MTho carried out the phylogenetic analyses, SNP mapping, RT-PCR, bisulfite sequencing and participated in sequence capture experiments. LP carried out the RNA-Seq experiments and its analysis. AP designed perl script to compute the frequency of CG dinucleotides. JD carried out the Sequence capture experiments. MThr carried out the syntenic analysis. SF and JPP participated in the design and coordination of the sequence capture, syntenic analysis and genetic mapping. CT designed and coordinated the study and carried out the evolution rate analysis. MTho, AVP and CT wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

MTho is supported by a Cifre PhD agreement n°817/2010 between the CNRS and the Biogemma Company. This work is supported by the “pole de competitivité Céréales Vallée”. CTa and AVP are supported by the CNRS, INSERM, Blaise Pascal and Auvergne Universities. AVP is supported by ANR “Dynam'Het” ANR-11 JSV2 009 01 and ANR “SINODYN” ANR-12-ISV6-

0001. CTa and AVP are supported by the Region Auvergne through “Life GRID” and a “Young Researcher Fellowship” respectively. We would like to thank J. Enjalbert for sharing results prior to publication, M. Abrouk and F. Choulet for critical reading and helpful suggestions and two anonymous reviewers for their fruitful comments.

References

1. Gaut BS: **Evolutionary dynamics of grass genomes**. *New Phytol* 2002, **154**:15–28.
2. Feldman M, Lupton F, Miller T: **Wheats**. In *Evol Crops Ed 2 Longman Sci Lond*. Edited by Smartt J, Simmonds N.; 1995:184–192.
3. Greer E, Martin AC, Pendle A, Colas I, Jones AME, Moore G, Shaw P: **The Ph1 Locus Suppresses Cdk2-Type Activity during Premeiosis and Meiosis in Wheat**. *PLANT CELL ONLINE* 2012, **24**:152–162.
4. Moore G, Devos KM, Wang Z, Gale MD: **Cereal Genome Evolution: Grasses, line up and form a circle**. *Curr Biol* 1995, **5**:737–739.
5. Yu J, Hu S, Wang J, Wong GK-S, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, et al.: **A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *indica*)**. *Science* 2002, **296**:79–92.
6. Goff SA, Ricke D, Lan T-H, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H, Hadley D, Hutchison D, Martin C, Katagiri F, Lange BM, Moughamer T, Xia Y, Budworth P, Zhong J, Miguel T, Paszkowski U, Zhang S, Colbert M, Sun W, Chen L, Cooper B, Park S, Wood TC, Mao L, Quail P, et al.: **A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*)**. *Science* 2002, **296**:92–100.
7. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberler G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otiillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, et al.: **The Sorghum bicolor genome and the diversification of grasses**. *Nature* 2009, **457**:551–556.
8. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, et al.: **The B73 Maize Genome: Complexity, Diversity, and Dynamics**. *Science* 2009, **326**:1112–1115.
9. Vogel JP, Garvin DF, Mockler TC, Schmutz J, Rokhsar D, Bevan MW, Barry K, Lucas S, Harmon-Smith M, Lail K, Tice H, (Leader) JS, Grimwood J, McKenzie N, Bevan MW, Huo N,

- Gu YQ, Lazo GR, Anderson OD, (Leader) JPV, You FM, Luo M-C, Dvorak J, Wright J, Febrer M, Bevan MW, Idziak D, Hasterok R, Garvin DF, Lindquist E, et al.: **Genome sequencing and analysis of the model grass *Brachypodium distachyon***. *Nature* 2010, **463**:763–768.
10. Consortium TIBGS: **A physical, genetic and functional sequence assembly of the barley genome**. *Nature* 2012, **491**:711–716.
11. Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C: **Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution**. *Plant Cell Online* 2008, **20**:11–24.
12. Akhunov ED, Akhunova AR, Linkiewicz AM, Dubcovsky J, Hummel D, Lazo G, Chao S, Anderson OD, David J, Qi L, Echalié B, Gill BS, Miftahudin, Gustafson JP, Rota ML, Sorrells ME, Zhang D, Nguyen HT, Kalavacharla V, Hossain K, Kianian SF, Peng J, Lapitan NLV, Wennerlind EJ, Nduati V, Anderson JA, Sidhu D, Gill KS, McGuire PE, Qualset CO, et al.: **Synteny perturbations between wheat homoeologous chromosomes caused by locus duplications and deletions correlate with recombination rates**. *Proc Natl Acad Sci* 2003, **100**:10836–10841.
13. Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier M-C, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C: **Megabase Level Sequencing Reveals Contrasted Organization and Evolution Patterns of the Wheat Gene and Transposable Element Spaces**. *Plant Cell Online* 2010, **22**:1686–1701.
14. Kellis M, Birren BW, Lander ES: **Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae***. *Nature* 2004, **428**:617–624.
15. Smet RD, Adams KL, Vandepoele K, Montagu MCEV, Maere S, Peer YV de: **Convergent gene loss following gene and genome duplications creates single-copy families in flowering plants**. *Proc Natl Acad Sci* 2013, **110**:2898–2903.
16. Schnable JC, Springer NM, Freeling M: **Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss**. *Proc Natl Acad Sci* 2011, **108**:4069–4074.
17. Cheng F, Wu J, Fang L, Sun S, Liu B, Lin K, Bonnema G, Wang X: **Biased Gene Fractionation and Dominant Gene Expression among the Subgenomes of *Brassica rapa***. *PLoS ONE* 2012, **7**:e36442.
18. Mayer KFX, Rogers J, Doležel J, Pozniak C, Eversole K, Feuillet C, Gill B, Friebe B, Lukaszewski AJ, Sourdille P, Endo TR, Kubaláková M, Číhalíková J, Dubská Z, Vrána J, Šperková R, Šimková H, Febrer M, Clissold L, McLay K, Singh K, Chhuneja P, Singh NK, Khurana J, Akhunov E, Choulet F, Alberti A, Barbe V, Wincker P, Kanamori H, et al.: **A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome**. *Science* 2014, **345**:1251788.
19. Groszmann M, Greaves IK, Fujimoto R, James Peacock W, Dennis ES: **The role of epigenetics in hybrid vigour**. *Trends Genet* 2013, **29**:684–690.
20. Jackson S, Chen ZJ: **Genomic and expression plasticity of polyploidy**. *Curr Opin Plant Biol* 2010, **13**:153–159.

21. Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, Ukomadu C, Sadler KC, Pradhan S, Pellegrini M, Jacobsen SE: **Conservation and divergence of methylation patterning in plants and animals.** *Proc Natl Acad Sci U S A* 2010, **107**:8689–8694.
22. Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O: **Selective epigenetic control of retrotransposition in Arabidopsis.** *Nature* 2009, **461**:427–430.
23. Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R: **Role of transposable elements in heterochromatin and epigenetic control.** *Nature* 2004, **430**:471–476.
24. Kinoshita Y, Saze H, Kinoshita T, Miura A, Soppe WJJ, Koornneef M, Kakutani T: **Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats.** *Plant J* 2006, **49**:38–45.
25. Melamed-Bessudo C, Levy AA: **Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis.** *Proc Natl Acad Sci* 2012, **109**:E981–E988.
26. Vongs A, Kakutani T, Martienssen RA, Richards EJ: **Arabidopsis thaliana DNA methylation mutants.** *Science* 1993, **260**:1926–1928.
27. Steward N, Kusano T, Sano H: **Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells.** *Nucleic Acids Res* 2000, **28**:3250–3259.
28. Teerawanichpan P, Chandrasekharan M, Jiang Y, Narangajavana J, Hall T: **Characterization of two rice DNA methyltransferase genes and RNAi-mediated reactivation of a silenced transgene in rice callus.** *Planta* 2004, **218**:337–349.
29. Fujimoto R, Sasaki T, Nishio T: **Characterization of DNA methyltransferase genes in Brassica rapa.** *Genes Genet Syst* 2006, **81**:235–242.
30. Brenchley R, Spannagl M, Pfeifer M, Barker GLA, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo M-C, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KFX, Edwards KJ, Bevan MW, Hall N: **Analysis of the bread wheat genome using whole-genome shotgun sequencing.** *Nature* 2012, **491**:705–710.
31. Murat F, Xu J-H, Tannier E, Abrouk M, Guilhot N, Pont C, Messing J, Salse J: **Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution.** *Genome Res* 2010, **20**:1545–1557.
32. Warren AS, Anandakrishnan R, Zhang L: **Functional bias in molecular evolution rate of Arabidopsis thaliana.** *BMC Evol Biol* 2010, **10**:125.
33. Yang Z: **PAML 4: Phylogenetic Analysis by Maximum Likelihood.** *Mol Biol Evol* 2007, **24**:1586–1591.

34. Duret L, Mouchiroud D: **Determinants of Substitution Rates in Mammalian Genes: Expression Pattern Affects Selection Intensity but Not Mutation Rate.** *Mol Biol Evol* 2000, **17**:68–070.
35. Zadoks JC, Chang TT, Konzak CF: **A decimal code for the growth stages of cereals.** *Weed Res* 1974, **14**:415–421.
36. Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F: **DNA Methylation Dynamics during Sexual Reproduction in *Arabidopsis thaliana*.** *Curr Biol* 2012, **22**:1825–1830.
37. Yamauchi T, Moritoh S, Johzuka-Hisatomi Y, Ono A, Terada R, Nakamura I, Iida S: **Alternative splicing of the rice OsMET1 genes encoding maintenance DNA methyltransferase.** *J Plant Physiol* 2008, **165**:1774–1782.
38. Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Gouis JL, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Safar J, Dolezel J, Rogers J, Vandepoele K, Aury J-M, Mayer K, Berges H, Quesneville H, Wincker P, et al.: **Structural and functional partitioning of bread wheat chromosome 3B.** *Science* 2014, **345**:1249721.
39. Genger RK, Kovac KA, Dennis ES, Peacock WJ, Finnegan EJ: **Multiple DNA methyltransferase genes in *Arabidopsis thaliana*.** *Plant Mol Biol* 1999, **41**:269–278.
40. Finnegan EJ, Kovac KA: **Plant DNA methyltransferases.** *Plant Mol Biol* 2000, **43**:189–201.
41. Chen M, Ha M, Lackey E, Wang J, Chen ZJ: **RNAi of met1 Reduces DNA Methylation and Induces Genome-Specific Changes in Gene Expression and Centromeric Small RNA Accumulation in *Arabidopsis* Allopolyploids.** *Genetics* 2008, **178**:1845–1858.
42. Diez CM, Roessler K, Gaut BS: **Epigenetics and plant genome evolution.** *Curr Opin Plant Biol* 2014, **18**:1–8.
43. Tran RK, Henikoff JG, Zilberman D, Ditt RF, Jacobsen SE, Henikoff S: **DNA Methylation Profiling Identifies CG Methylation Clusters in *Arabidopsis* Genes.** *Curr Biol* 2005, **15**:154–159.
44. Lynch M, Conery JS: **The Evolutionary Fate and Consequences of Duplicate Genes.** *Science* 2000, **290**:1151–1155.
45. Yang L, Gaut BS: **Factors that Contribute to Variation in Evolutionary Rate among *Arabidopsis* Genes.** *Mol Biol Evol* 2011, **28**:2359–2369.
46. Yang L, Takuno S, Waters ER, Gaut BS: **Lowly Expressed Genes in *Arabidopsis thaliana* Bear the Signature of Possible Pseudogenization by Promoter Degradation.** *Mol Biol Evol* 2011, **28**:1193–1203.
47. Duret L, Galtier N: **Biased gene conversion and the evolution of mammalian genomic landscapes.** *Annu Rev Genomics Hum Genet* 2009, **10**:285–311.
48. Smith ZD, Meissner A: **DNA methylation: roles in mammalian development.** *Nat Rev Genet* 2013, **14**:204–220.

49. Ponger L, Duret L, Mouchiroud D: **Determinants of CpG Islands: Expression in Early Embryo and Isochore Structure.** *Genome Res* 2001, **11**:1854–1860.
50. Suzuki S, Shaw G, Kaneko-Ishino T, Ishino F, Renfree MB: **The Evolution of Mammalian Genomic Imprinting Was Accompanied by the Acquisition of Novel CpG Islands.** *Genome Biol Evol* 2011, **3**:1276–1283.
51. Serres-Giardi L, Belkhir K, David J, Glémin S: **Patterns and Evolution of Nucleotide Landscapes in Seed Plants.** *Plant Cell Online* 2012, **24**:1379–1397.
52. Saintenac C, Falque M, Martin OC, Paux E, Feuillet C, Sourdille P: **Detailed Recombination Studies Along Chromosome 3B Provide New Insights on Crossover Distribution in Wheat (*Triticum aestivum* L.).** *Genetics* 2009, **181**:393–403.
53. Yamauchi T, Johzuka-Hisatomi Y, Terada R, Nakamura I, Iida S: **The MET1b gene encoding a maintenance DNA methyltransferase is indispensable for normal development in rice.** *Plant Mol Biol* 2014:1–14.
54. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA: **Direct selection of human genomic loci by microarray hybridization.** *Nat Methods* 2007, **4**:903–905.
55. Saintenac C, Jiang D, Akhunov ED: **Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome.** *Genome Biol* 2011, **12**:R88.
56. Gervais L, Dedryver F, Morlais J-Y, Bodusseau V, Negre S, Bilous M, Groos C, Trottet M: **Mapping of quantitative trait loci for field resistance to *Fusarium* head blight in an European winter wheat.** *Theor Appl Genet* 2003, **106**:961–970.
57. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792–1797.
58. Castresana J: **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.** *Mol Biol Evol* 2000, **17**:540–552.
59. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O: **New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.** *Syst Biol* 2010, **59**:307–321.
60. Letunic I, Bork P: **Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy.** *Nucleic Acids Res* 2011, **39**(suppl 2):W475–W478.
61. Salamov AA, Solovyev VV: **Ab initio Gene Finding in *Drosophila* Genomic DNA.** *Genome Res* 2000, **10**:516–522.
62. Rice P, Longden I, Bleasby A: **EMBOSS: The European Molecular Biology Open Software Suite.** *Trends Genet* 2000, **16**:276–277.
63. Rustenholz C, Choulet F, Laugier C, Šafář J, Šimková H, Doležel J, Magni F, Scalabrin S, Cattonaro F, Vautrin S, Bellec A, Bergès H, Feuillet C, Paux E: **A 3,000-Loci Transcription Map of Chromosome 3B Unravels the Structural and Functional Features of Gene Islands in Hexaploid Wheat.** *Plant Physiol* 2011, **157**:1596–1608.

64. Trapnell C, Pachter L, Salzberg SL: **TopHat: discovering splice junctions with RNA-Seq.** *Bioinformatics* 2009, **25**:1105–1111.
65. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**:R25.
66. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L: **Improving RNA-Seq expression estimates by correcting for fragment bias.** *Genome Biol* 2011, **12**:R22.
67. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Methods* 2008, **5**:621–628.
68. Bogorad L, Gubbins EJ, Krebbers ET, Larrinua IM, Mulligan BJ, Muskavitch KMT, Orr EA, Rodermeil SR, Schantz R, Steinmetz AA, De Vos G, Ye YK: **Cloning and physical mapping of maize plastid genes.** *Methods Enzymol* 1983, **97**:524–554.
69. Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M: **Identification and validation of reference genes for quantitative RT-PCR normalization in wheat.** *BMC Mol Biol* 2009, **10**:11.
70. Gruntman E, Qi Y, Slotkin RK, Roeder T, Martienssen RA, Sachidanandam R: **Kismeth: Analyzer of plant methylation states through bisulfite sequencing.** *BMC Bioinformatics* 2008, **9**:371–371.
71. Khan A, Enjalbert J, Marsollier A-C, Rousselet A, Goldringer I, Vitte C: **Vernalization treatment induces site-specific DNA hypermethylation at the VERNALIZATION-A1 (VRN-A1) locus in hexaploid winter wheat.** *BMC Plant Biol* 2013, **13**:209.

Figure legends

Figure 1 Gene structures of *TaMET1* genes. A) The two *MET1* genes from rice (*OsMET1a*; [GenBank: Os03g58400] and *OsMET1b*; [GenBank: Os07g08500]) were used as a reference to define two *MET1* lineages (hereafter called MET-1a and MET-1b lineages). *OsMET1a* and *OsMET1b* are shown at the top of each lineage. Gene structures and splice junctions organize *TaMET1* genes in 11 exons (black and grey boxes) and 10 introns (horizontal lines). The three distinct protein domains identified are indicated at the top of the figure: DNMT1-RFD (light grey), BAH-domain (dark grey) and DNA methyltransferase (black). *TaMET1* genes from chromosome 2 and 7 as well as chromosome 5B are predicted to yield a full-length *MET1* protein while 5A and 5D contain stop codons (*) and deletions (Δ). 5' and 3'UTRs are indicated as smaller boxes. Potential promoter regions were defined as the ~2-3 kb region upstream of the

coding sequences and are indicated as a thicker blue line. Unknown sequence insertion (*TaMET-7DI*) is indicated as a red line and transposable element insertion of Stowaway, Gypsy, Mariner and Mutator are indicated as red boxes. **B)** Combination of methods used in this study.

Figure 2 *MET1* phylogenetic trees in flowering plants and syntenic conservation. A)

Maximum likelihood (ML) phylogenetic tree. The ML tree was inferred using cDNA sequences from dicot species including *Arabidopsis thaliana* (*AtMET1*: [TAIR:At5g49160]), *Brassica rapa* (*BrMET1a* and *BrMET1b* respectively [GenBank: AB251938 and AB25937]), *Pisum sativum* (*PsMET1*: [GenBank: AF034419]), *Medicago truncatula*: (*MtMET1*: [Phytozome: Medtr6g065580]), *Daucus carota* (*DcMET1*: [GenBank: AF007807]), *Nicotiana tabacum* (*NtMET1*: [Genbank: AB030726]), *Solanum lycopersicum* (*SlMET1*: [GenBank: AJ002140]), and monocot species including *Sorghum bicolor* (*SbMET1a* and *SbMET1b* respectively [Phytozome: Sb01g005084 and Sb02g004680]), *Zea mays* (*ZmMET1b-1* and *ZmMET1b-2* respectively [Phytozome: GRMZM2G333916 and GRMZM2G334041]), *Oriza sativa* (*OsMET1a* and *OsMET1b* respectively: [Genbank AB362510 and AB362511]), *Brachypodium distachyon* (*BdMET1a* and *BdMET1b* respectively [Phytozome: Bradi1g05380 and Bradi1g55287]), *Hordeum vulgare* (*HvMET1-2* and *HvMET1-7* respectively: [EnsemblGenomes: MLOC_61904.6 and: MLOC_10988.2]), *HvMET1-5* predicted from [GenBank: CAJW010043285]) and *Triticum aestivum*: *TaMET-2A1*, *-2B1*, *-2D1*, *-5A1*, *5B1*, *5D1*, *7A1*, *7B1* and *7DI* (this study from IWGSC contigs). Numbers above branches indicate bootstrap values. Two monophyletic groups were defined and called the MET-1a and MET-1b lineages according to the *OsMET1a* and *OsMET1b* genes (indicated in boxes). The two gene duplication events observed in wheat and barley are indicated by an arrow. **B) Micro-syteny analysis.** Micro-syteny was established by BLASTn analysis between genes from brachypodium, rice and IWGSC sequences from wheat chromosomes as described in Methods. Syntenic conservation between rice (in white) and brachypodium (in black) correspond to the number of conserved genes in wheat / number of genes

in the syntenic region from brachypodium or rice expressed as percentage. **C) Gene structures of *MET1* genes at chromosome 5.** BLASTn analysis was used to align *MET1* genes from *Hordeum vulgare* (*HvMET1-5*: [GenBank: CAJW010043285]), *Triticum urartu* (*TuMET1-5*: scaffold15783), *Aegilops tauschii* (*AetMET1-5*: scaffold164515) and *Triticum aestivum* (*TaMET-5A1* and *TaMET-5D1*) (this study).

Figure 3 Evolution rate analyses of *TaMET1* genes. A) Pair-wise comparisons. Synonymous (dS), non-synonymous (dN) and evolution rate (ω) are expressed in substitution / site and were performed using *Codeml*. *TaMET1* from wheat homoeologous group 2 (grey boxes) were compared to *Brachypodium distachyon* [Phytozome: Bradi1g05380 and Bradi1g55287], *Sorghum bicolor* [Phytozome: Sb01g005084 and Sb02g004680], *Oriza sativa* [Genbank AB362510 and AB362511], *Zea mays* [Phytozome: GRMZM2G333916 and GRMZM2G334041], *Hordeum vulgare* [EnsemblGenomes: MLOC_61904.6 and: MLOC_10988.2 and GenBank: CAJW010043285] and *Triticum aestivum* (*TaMET-2A1*, *2B1*, *-2D1*, *5A1*, *5B1*, *5D1*, *7A1*, *7B1* and *7D1*) (accession numbers in additional file 8). Mean values of dS, dN and ω (with $\omega = dN/dS$) were then computed for each homoeologous group. Whiskers represent the 10-90% range of mean values, boxes represent interquartile distances, the horizontal line across whiskers represents the median, and “+” the mean values. Kruskal Wallis non-parametric tests were applied to determine significant differences between mean values (*: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$). **B) PAML branch model.** Tree topology was defined by a protein alignment using the same monocot species as in A). The two monophyletic groups MET-1a and MET-1b are indicated at the right. Three distinct evolution rates ω_0 (red branches), ω_1 (black branches) and ω_2 (blue branches) are indicated as well as the two gene duplication events described in figure 1 (arrow).

Figure 4 Expression profiles of wheat *TaMET1*. A) Wheat RNA-seq. Data are expressed in Fragment Reads per Kilobase of Exon Model (FPKM) for each homoeologous group. FPKM were

computed according to the following formula: $FPKM = 10^9 \times (C/NL)$ where C is the number of mappable reads on a feature, N is the total number of reads in the experiment and L (length) is the sum of exonic sequences in base pairs. Genes were considered to be expressed only for FPKM values > 0.15 . Samples covering various tissues (root, leaf, stem, spike and grain) and developmental stages indicated as Zadoks (Z) scales are indicated at the bottom. Expression levels from *TaMET1* genes of a given homoeologous group were used to compute mean values. *TaMET1* from group 2 are expressed at higher levels in comparison to the other *MET1* copies and display two main peaks of expression in the stem (Z30) and spike (Z32). **B) RT-PCR gel analysis.** RT-PCR with (+) or without (-) reverse transcriptase was performed for all *TaMET1* genes. Control PCR reactions were performed for two constitutively expressed genes corresponding to Ta4045 and Ta54227 selected according to Paolacci *et al.* 2009. **C) Quantitative RT PCR.** Reactions were performed for *TaMET-2A1*, *2B1* and *2D1*. Expression is relative to (dark grey) and Ta54227 (light grey). **D) Schematic representation of *TaMET1* loci.** Position of *MET1* genes on chromosome maps of 2B, 5B and 7A (chromosome arms in grey and centromere as white ellipse, not to scale). Top and bottom positions on ITMI reference map are indicated in cM.

Figure 5 CG enrichment at potential promoters of *TaMET1* genes from homoeologous group 5 and 7. A) Frequency of CG dinucleotides. Frequencies were computed every 50bp and are shown for each homoeologous group. Putative promoter region and coding sequence are delimited by respectively a blue and white box. Black bars numbered from PCR1 to PCR4 highlight the four regions studied by bisulfite sequencing and are indicated above the graphs. Region 4 is specific to the 7D homoeolog. Arrows indicate the putative transcription start site according to the RNA-Seq data. **B) Mean values of CG dinucleotides.** Mean values of the number of CG dinucleotides of the three homoeologs (A, B and D) for a given homoeologous group (2, 5 and 7) in the putative

promoter (blue) and coding (white) sequence regions. Differences between groups 5 and 7 putative promoter regions and group 2 are indicated above the histogram. Statistical significance was confirmed with a Kruskal Wallis non parametric tests with *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$. **C) DNA methylation profiles as determined by bisulfite sequencing.** Percentages of methylated cytosines of the four amplicons (PCR1 to 4) displayed in figure 6A were determined after bisulfite sequencing. Percentages of methylation were recorded at each cytosine position and were used to compute a mean value for each amplicon in the CG, CHG or CHH sequence contexts.

Figure 6 Scheme of the genome dynamics at *TaMET1* loci in the course of evolution. Scheme of the emergence of the nine *TaMET1* genes is given at the left. Evolution time is given from the top to the bottom in MYa. The MET-1a lineage is expressed (green) and evolved at a low evolution rate while genes in the MET-1b lineage are repressed (red) or evolved as pseudogenes (Ψ , black boxes). The hypothetical succession of genomic events occurring at or including *TaMET1* is proposed at the right.

Additional files

Additional file 1 Likelihood ratio tests (LRT). A) Likelihood ratio test (LRT). LRT has been used to compare 11 hypotheses (H_{1-11}) in respect to the null hypothesis (H_0) in which all the branches have the same evolution rate (ω_0). Hypotheses were designed to test if the MET1 phylogenetic tree includes up to three evolution rates (ω_0 , ω_1 and ω_2) and if these variations in ω values are long term changes (*i.e* if all the branches below a duplication event display the same ω value) or increase only after a duplication event and then is relaxed to ω_0 . **B) Details of the 11 hypotheses tested in the branch model described in Fig 3B.** For each hypothesis tested, a tree

file in Newick format and a graphic representation highlighting the branches considered in the tested hypothesis are given.

Additional file 2 Genetic positions of *TaMET1* loci. Distal and proximal markers from the ITMI reference map and flanking the 2B, 5B and 7A *TaMET1* loci are given in cM.

Additional file 3 Alignment at putative promoter regions of *TaMET1* genes from homoeologous group 7. *Hordeum vulgare* chromosome 7 [EnsemblGenomes: MLOC_10988.2], *Triticum aestivum* chromosome 7A [IWGSC: 7AL:4532056], 7B [IWGSC: 7BL:6682174] and 7D [IWGSC: 7DL:3392185], *Triticum urartu* chromosome 7 [EnsemblGenomes: scaffold38640], *Triticum tauschii* chromosome 7 [EnsemblGenomes: scaffold2203], Alignment were performed with *MUSCLE* and refined by *jalview*.

Additional file 4 Bisulfite analysis of putative promoter region of homoeologous group 5. A) Frequencies of CG dinucleotides were computed every 50bp of the putative promoter regions of homoeologous group 5. 5A (black), 5B (white) and 5D (grey). Black bars numbered from 1 to 4 highlight the two regions studied by bisulfite sequencing. **B)** Kismeth outputs of the percentage of methylated cytosines in CG (red), CHG (green) and CHH (blue) context.

Additional file 5 Bisulfite analysis of putative promoter region of homoeologous group 7. A) Frequencies of CG dinucleotides were computed every 50bp of the putative promoter regions of homoeologous group 7. 7A (black), 7B (white) and 7D (grey). Black bars numbered from 1 to 4 highlight the two regions studied by bisulfite sequencing. **B)** Kismeth outputs of the percentage of methylated cytosines in CG (red), CHG (green) and CHH (blue) context.

Additional file 6 Primers used in RT-PCR, RT-qPCR, mapping and bisulfite experiments.

Additional file 7 Virtual physical map reconstruction at *TaMET1* loci from micro-synteny data. Physical maps for Os and Bd, virtual physical map based on IWGSC surveys organized from rice and brachypodium orthologs. *TaMET1* loci are highlighted in yellow. Note that two overlapping contigs were found at *TaMET-5A1* indicating that these two IWGSC contigs were not assembled together in the course of the assembly process.

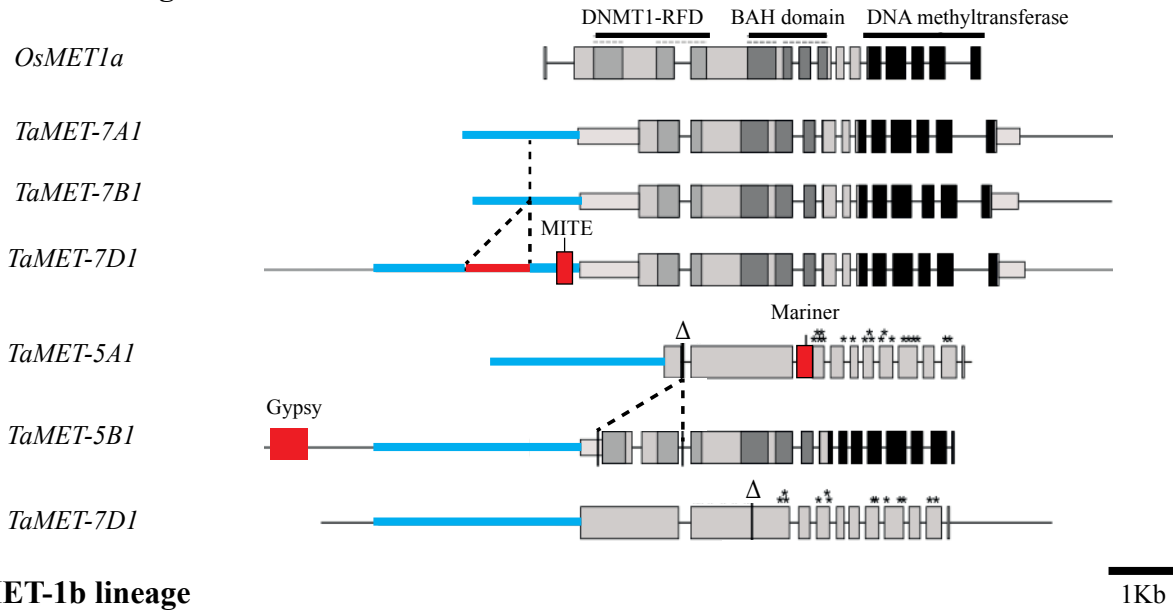
Additional file 8 RNA-Seq samples used in this study.

Additional file 9 Controls in bisulfite experiments. A) Methylation rates at two VRN-A1 regions called 0.0 k and 9.2 k (adapted from [69]). **B)** Structure of the *VRN-A1* gene. **C)** Typical results from bisulfite experiments for 0.0 k (no CG methylation) and 9.2 k (high CG methylation).

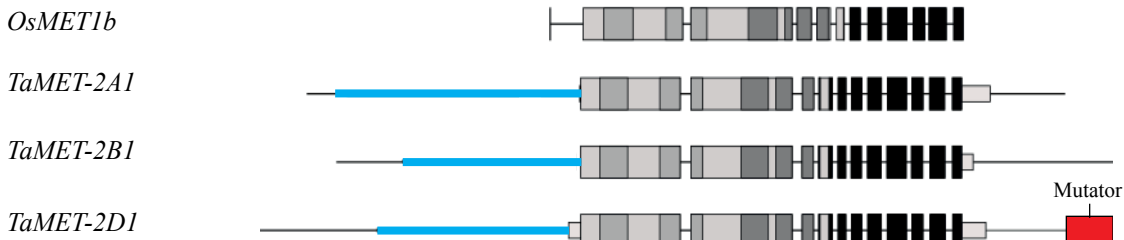
Figure 1

A

MET-1a lineage



MET-1b lineage



B

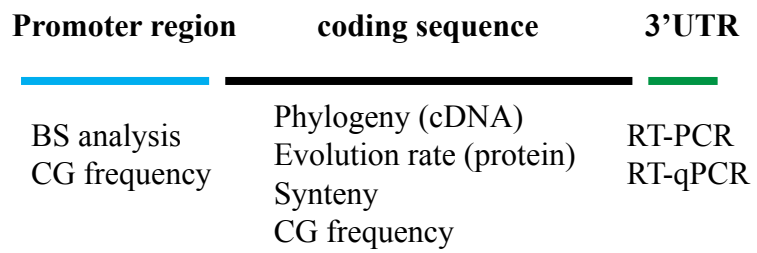
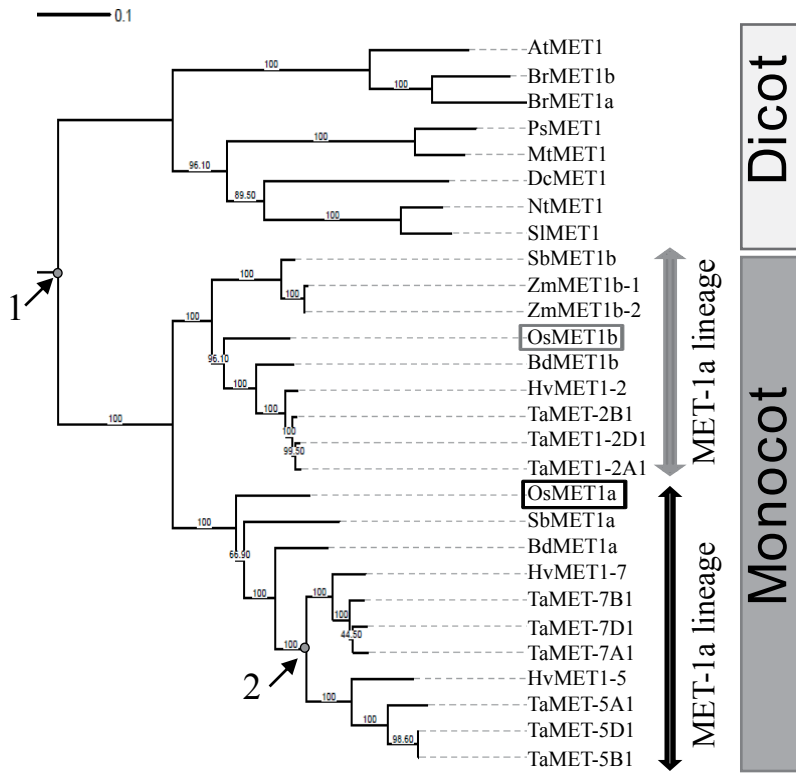
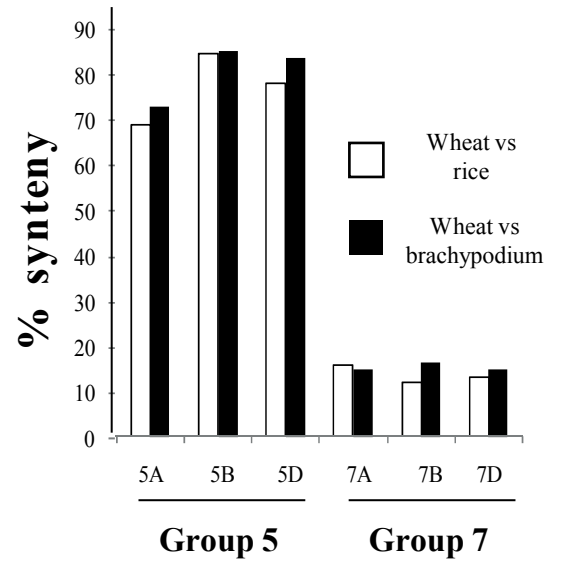


Figure 2

A



B



C

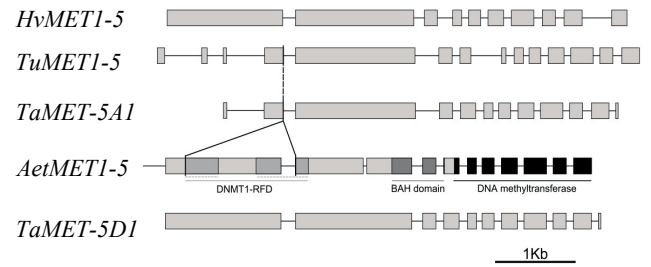
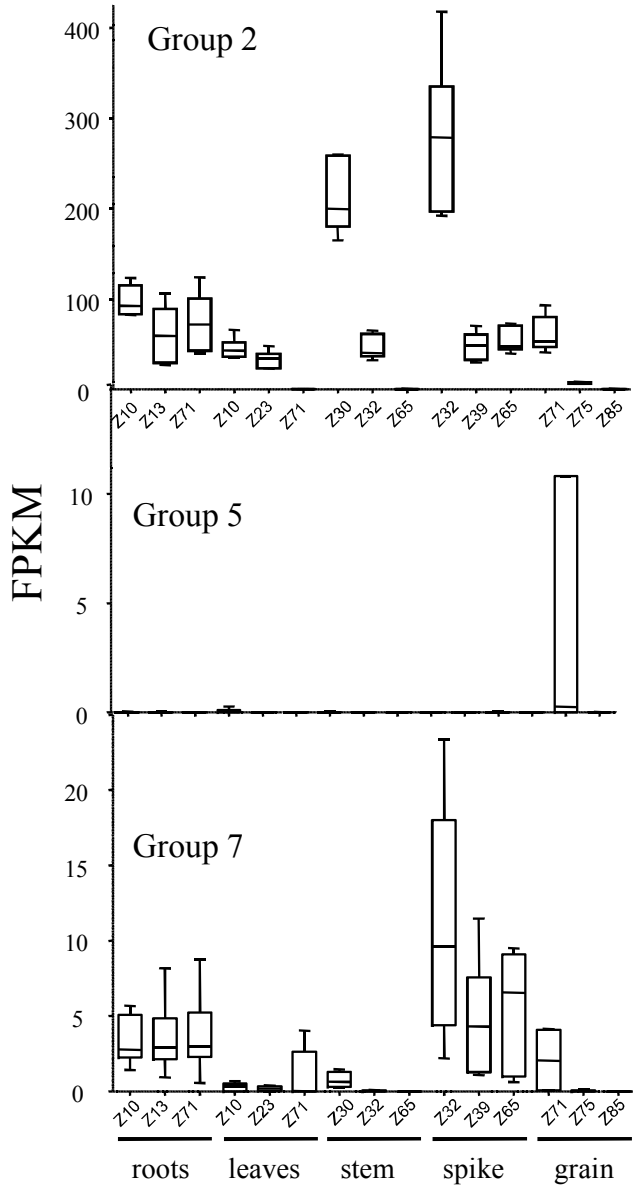
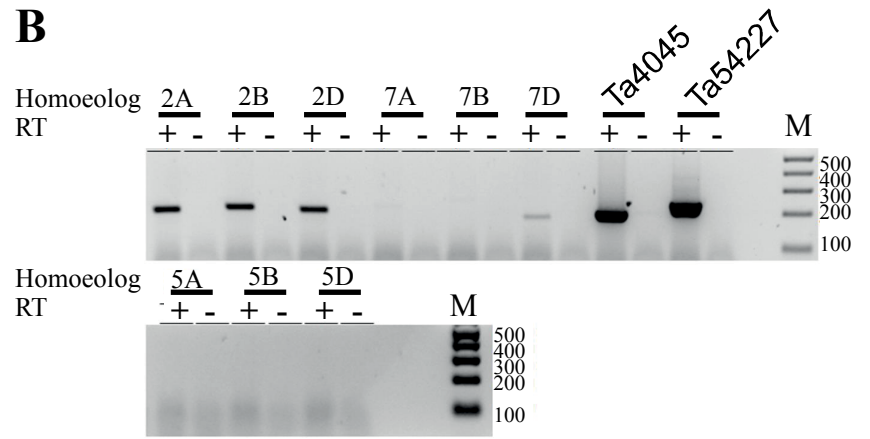


Figure 4

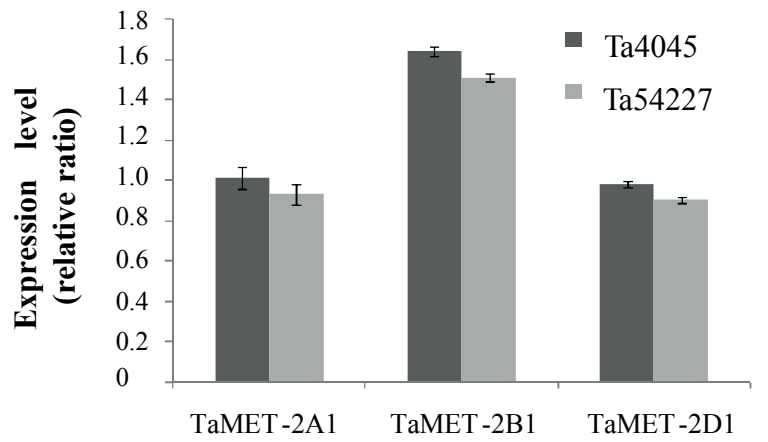
A



B



C



D

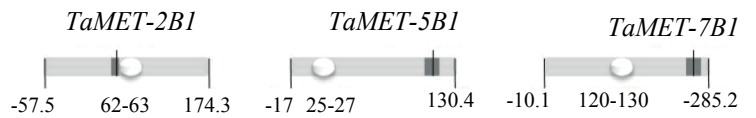


Figure 5

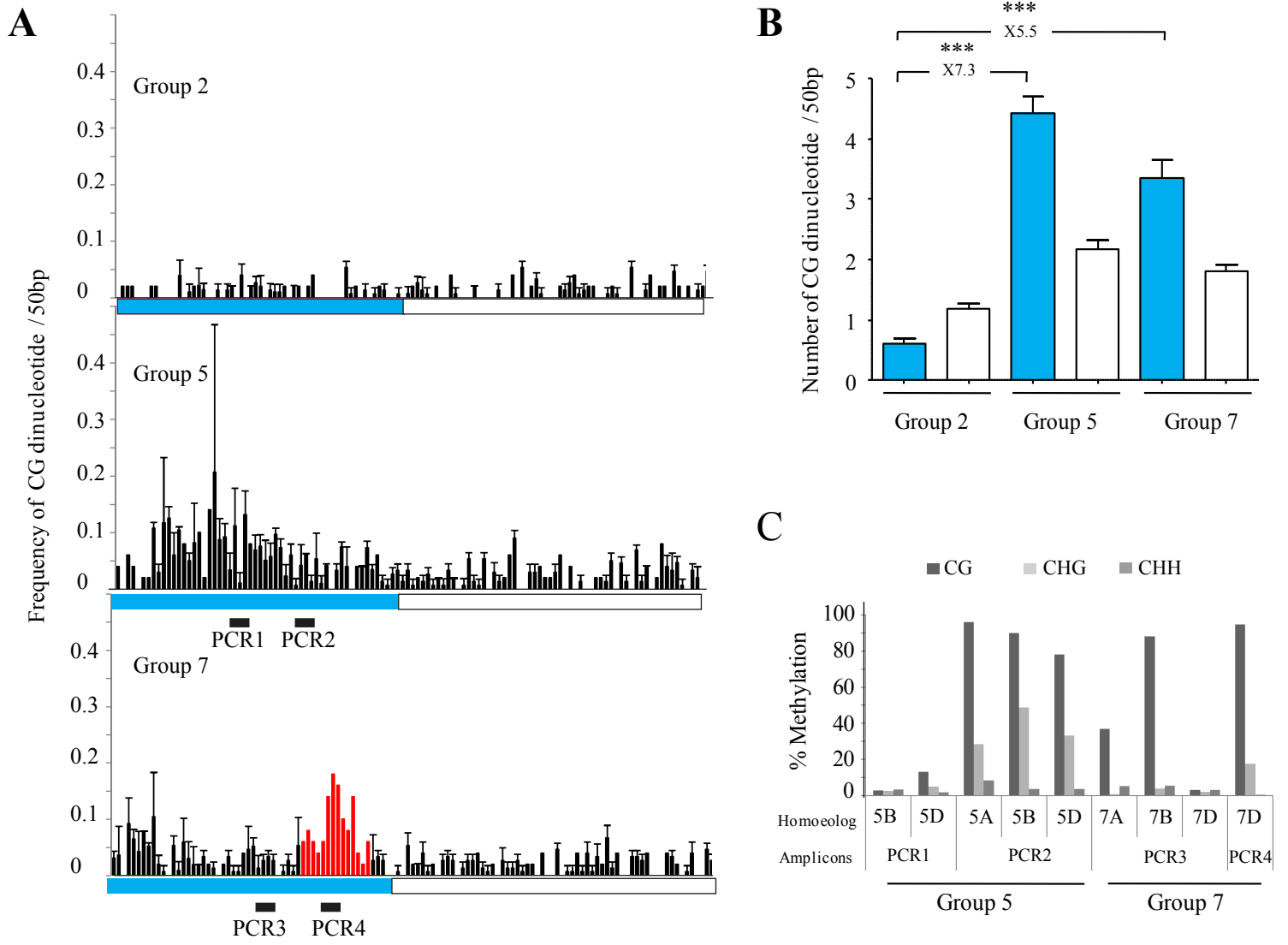
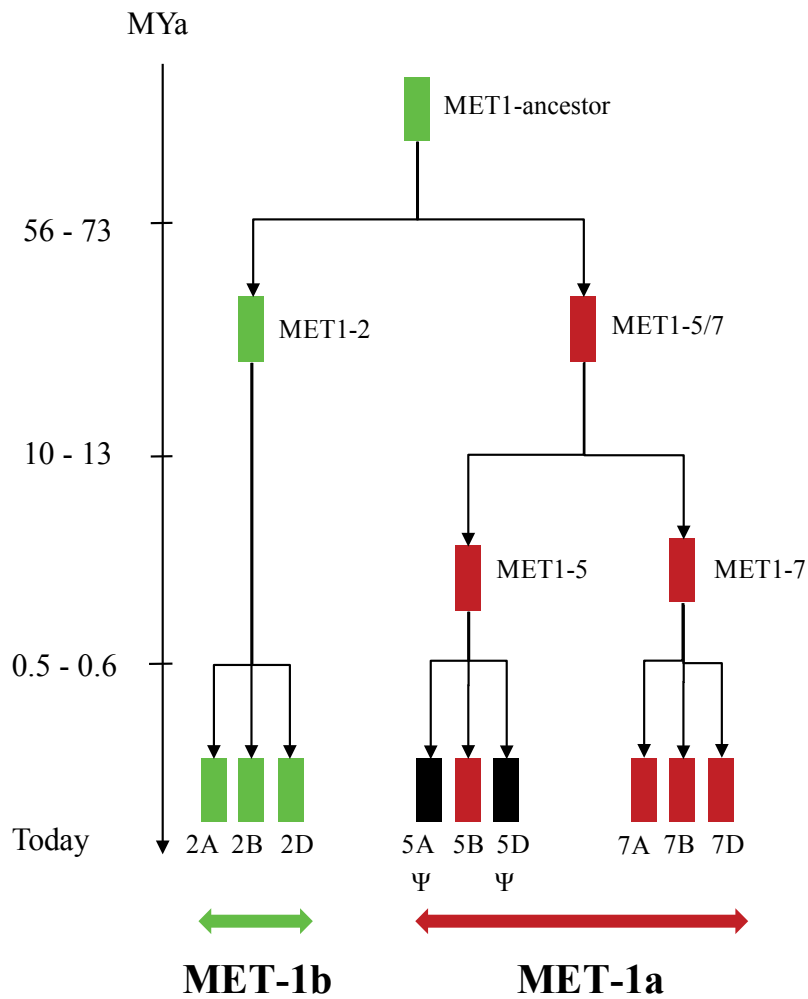
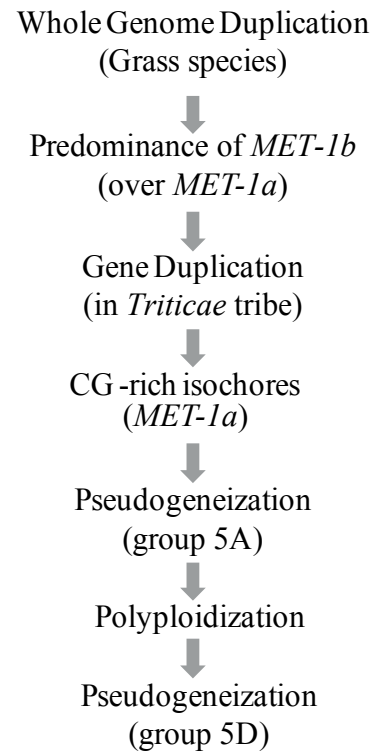


Figure 6



Genomic events



Additional file 1

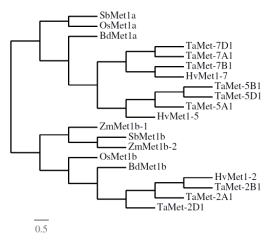
Additional file 1A: Likelihood ratio tests (LRT)

Hypothesis	<i>Lnl</i>	$2\Delta l$	<i>ddl</i>	LRT	ω_0	ω_1	ω_2	P<0.05	P<0.01
h0	-21434.8	-	-	-	0.23993	-	-	-	-
h1	-21330.6	104.1	1	0	0.17544	0.55407	-	***	***
h2	-21434.3	0.4768	1	0.0383	0.23822	0.30632	-	*	ns
h3	-21434.0	0.7571	1	0.0383	0.23567	0.26794	-	*	ns
h4	-21434.8	0.0009	1	1	0.23986	0.24225	-	ns	ns
h5	-21362.2	72.6	1	0	0.12974	0.32857	-	***	***
h6	-21434.8	-0.00004	1	1	0.23993	0.36697	-	ns	ns
h7	-21362.2	72.6	1	0	0.12974	0.32858	-	***	***
h8	-21434.6	0.2523	1	0.0617	0.21638	0.2417	-	ns	ns
h9	-21433.4	1.4	1	0.0235	0.22161	0.2512	-	*	ns
h10	-21330.6	104.1	1	0	0.17544	0.55407	-	***	***
h11	-21326.0	108.8	2	0	0.12193	0.21988	0.48706	***	***

Additional file 1B: Details of the 11 hypotheses tested in the branch model described in Fig 3B

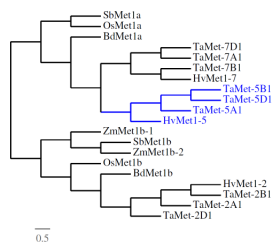
h0: all the branches evolve at the same evolution rate

(((((TaMet-2D1,(TaMet-2A1,(TaMet-2B1,HvMet1-2))),BdMet1b),OsMet1b),((ZmMet1b-2,SbMet1b),ZmMet1b-1)),(((HvMet1-5,(TaMet-5A1,(TaMet-5D1,TaMet-5B1))),((HvMet1-7,TaMet-7B1),(TaMet-7A1,TaMet-7D1))),BdMet1a),(OsMet1a,SbMet1a)));



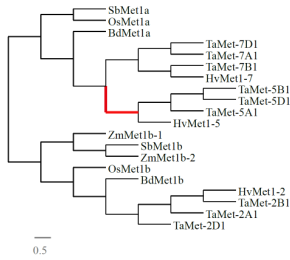
h1: the branches including group 5 homoeologs and HvMet1-5 evolve at the same rate which is significantly different from the rest of the tree

(((((TaMet-2D1,(TaMet-2A1,(TaMet-2B1,HvMet1-2))),BdMet1b),OsMet1b),((ZmMet1b-2,SbMet1b),ZmMet1b-1)),(((HvMet1-5 #1,(TaMet-5A1 #1,(TaMet-5D1 #1,TaMet-5B1 #1) #1) #1),((HvMet1-7,TaMet-7B1),(TaMet-7A1,TaMet-7D1))),BdMet1a),(OsMet1a,SbMet1a)));



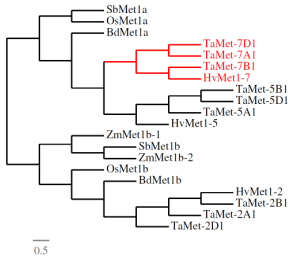
h2: evolution rate increases soon after the second duplication and return to its initial rate

(((TaMet-2D1,(TaMet-2A1,(TaMet-2B1,HvMet1-2))),BdMet1b),OsMet1b),((ZmMet1b-2,SbMet1b),ZmMet1b-1)),(((HvMet1-5,(TaMet-5A1,(TaMet-5D1,TaMet-5B1)))#1, ((HvMet1-7,TaMet-7B1),(TaMet-7A1,TaMet-7D1))),BdMet1a),(OsMet1a,SbMet1a));



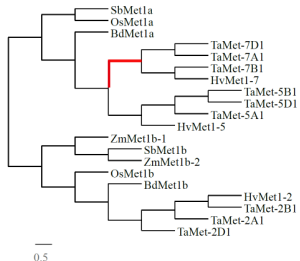
h3: the branches including group 7 homoeologs and HvMet1-7 evolve at the same rate which is significantly different from the rest of the tree

(((TaMet-2D1,(TaMet-2A1,(TaMet-2B1,HvMet1-2))),BdMet1b),OsMet1b),((ZmMet1b-2,SbMet1b),ZmMet1b-1)),(((HvMet1-5,(TaMet-5A1,(TaMet-5D1,TaMet-5B1 #1))),((HvMet1-7 #1,TaMet-7B1 #1) #1,(TaMet-7A1 #1,TaMet-7D1 #1) #1) #1),BdMet1a),(OsMet1a,SbMet1a));



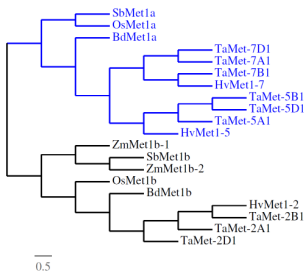
h4: evolution rate increases soon after the second duplication and return to its initial rate

(((TaMet-2D1,(TaMet-2A1,(TaMet-2B1,HvMet1-2))),BdMet1b),OsMet1b),((ZmMet1b-2,SbMet1b),ZmMet1b-1)),(((HvMet1-5,(TaMet-5A1,(TaMet-5D1,TaMet-5B1 #1))),((HvMet1-7,TaMet-7B1),(TaMet-7A1,TaMet-7D1) #1),BdMet1a),(OsMet1a,SbMet1a));

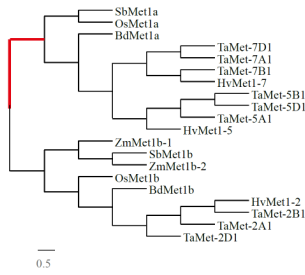


h5: the branches including MET1a lineage evolve at the same rate which is significantly different from the rest of the tree (i.e from MET1b lineage)

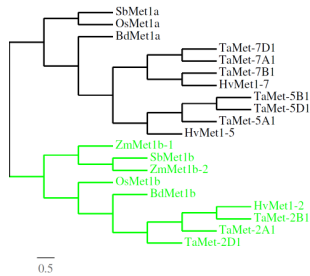
(((TaMet-2D1,(TaMet-2A1,(TaMet-2B1,HvMet1-2))),BdMet1b),OsMet1b),((ZmMet1b-2,SbMet1b),ZmMet1b-1)),(((HvMet1-5 #1,(TaMet-5A1 #1,(TaMet-5D1 #1,TaMet-5B1 #1) #1) #1,((HvMet1-7 #1,TaMet-7B1 #1) #1,(TaMet-7A1 #1,TaMet-7D1 #1) #1) #1),BdMet1a #1),(OsMet1a #1,SbMet1a #1) #1);



h6: evolution rate increases soon after the first duplication and return to its initial rate
 (((((TaMet-2D1,(TaMet-2A1,(TaMet-2B1,HvMet1-2))),BdMet1b),OsMet1b),((ZmMet1b-2,SbMet1b),ZmMet1b-1)),(((HvMet1-5,(TaMet-5A1,(TaMet-5D1,TaMet-5B1))),((HvMet1-7,TaMet-7B1),(TaMet-7A1,TaMet-7D1))),BdMet1a),(OsMet1a,SbMet1a)) #1);

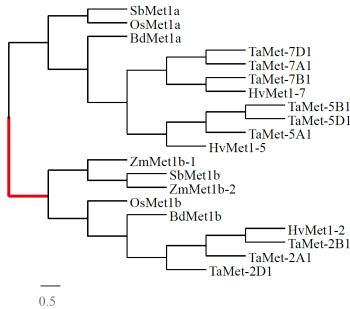


h7: the branches including MET1b lineage evolve at the same rate which is significantly different from the rest of the tree (i.e. from MET1a lineage)
 (((((TaMet-2D1 #1,(TaMet-2A1 #1,(TaMet-2B1 #1,HvMet1-2 #1) #1) #1),BdMet1b #1),OsMet1b#1),((ZmMet1b-2 #1,SbMet1b #1),ZmMet1b-1 #1) #1),(((HvMet1-5,(TaMet-5A1,(TaMet-5D1,TaMet-5B1))),((HvMet1-7,TaMet-7B1),(TaMet-7A1,TaMet-7D1))),BdMet1a),(OsMet1a,SbMet1a)));



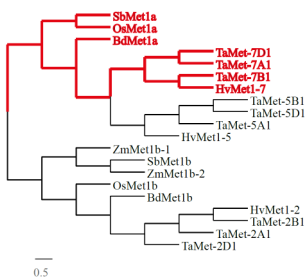
h8: evolution rate increases soon after the first duplication and return to its initial rate

(((TaMet-2D1,(TaMet-2A1,(TaMet-2B1,HvMet1-2))),BdMet1b),OsMet1b),((ZmMet1b-2,SbMet1b),ZmMet1b-1) #1,(((HvMet1-5,(TaMet-5A1,(TaMet-5D1,TaMet-5B1))),((HvMet1-7,TaMet-7B1),(TaMet-7A1,TaMet-7D1))),BdMet1a),(OsMet1a,SbMet1a)));



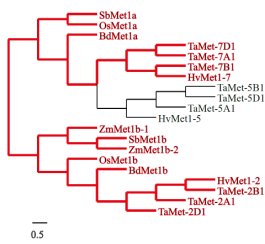
h9: the branches including MET1a lineage except group 5 homoeologs and HvMet1-5 evolve at the same rate which is significantly different from the rest of the tree

(((TaMet-2D1,(TaMet-2A1,(TaMet-2B1,HvMet1-2))),BdMet1b),OsMet1b),((ZmMet1b-2,SbMet1b),ZmMet1b-1),(((HvMet1-5,(TaMet-5A1,(TaMet-5D1,TaMet-5B1))),((HvMet1-7 #1,TaMet-7B1 #1) #1,(TaMet-7A1 #1,TaMet-7D1 #1) #1) #1),BdMet1a #1),(OsMet1a #1,SbMet1a #1) #1);



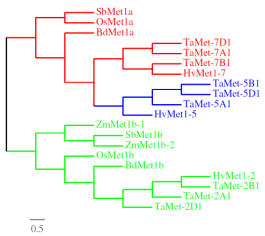
h10: all branches except group 5 homoeologs and HvMet1-5 evolve at a different rate which is significantly different from group 5 homoeologs and HvMet1-5 (reciprocal to h1)

(((TaMet-2D1 #1,(TaMet-2A1 #1,(TaMet-2B1 #1,HvMet1-2 #1) #1) #1,BdMet1b #1) #1,OsMet1b#1) #1,((ZmMet1b-2 #1,SbMet1b #1) #1,ZmMet1b-1 #1) #1,(((HvMet1-5,(TaMet-5A1,(TaMet-5D1,TaMet-5B1))) ,((HvMet1-7 #1,TaMet-7B1 #1) #1,(TaMet-7A1 #1,TaMet-7D1 #1) #1) #1) #1,BdMet1a #1) #1,(OsMet1a #1,SbMet1a #1) #1) #1);



h11: 3 evolution rates are observed, one for branches including group 5 homoeologs and HvMet1-5, one for group 7 homoeologs and HvMet1-7 and one for the MET1b lineage

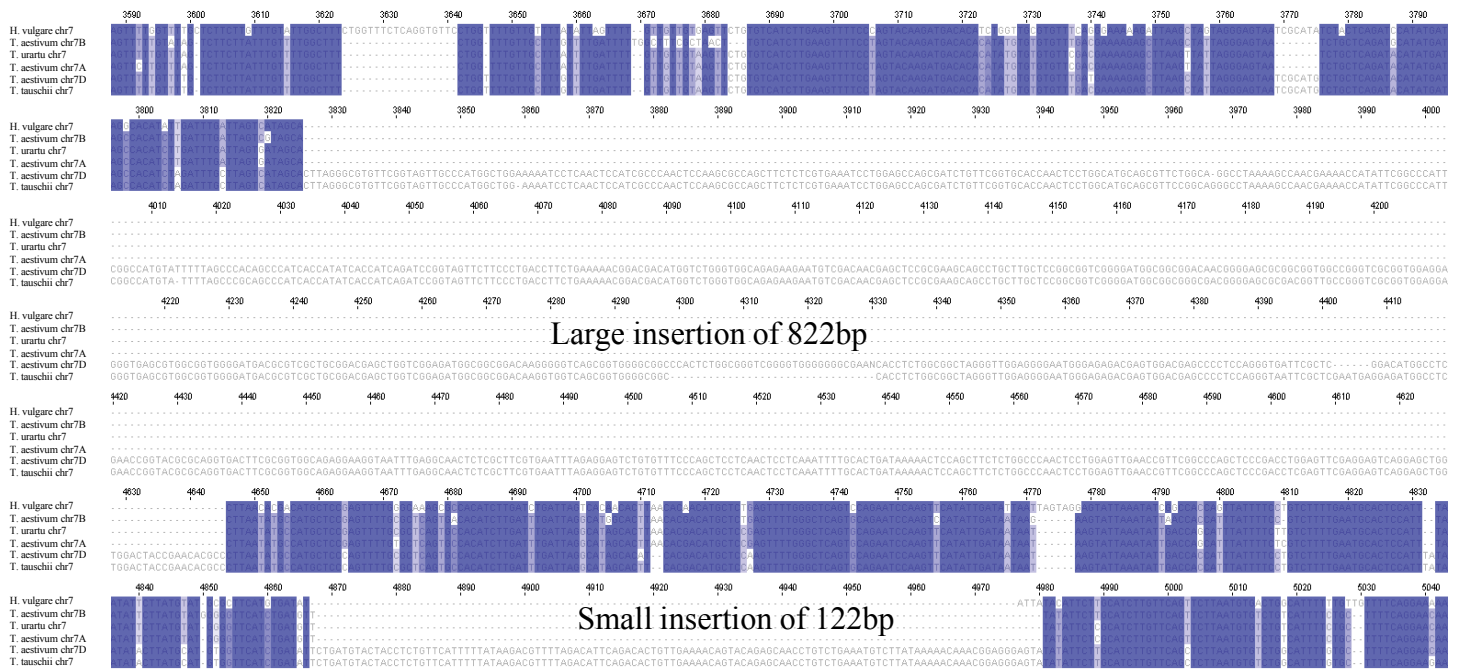
(((TaMet-2D1 #2,(TaMet-2A1 #2,(TaMet-2B1 #2,HvMet1-2 #2) #2) #2,BdMet1b #2),OsMet1b #2) #2,((ZmMet1b-2 #2,SbMet1b #2) #2,ZmMet1b-1 #2) #2,(((HvMet1-5,(TaMet-5A1,(TaMet-5D1,TaMet-5B1))) ,((HvMet1-7 #1,TaMet-7B1 #1) #1,(TaMet-7A1 #1,TaMet-7D1 #1) #1) #1) #1) #1,BdMet1a #1) #1,(OsMet1a #1,SbMet1a #1) #1) #1);



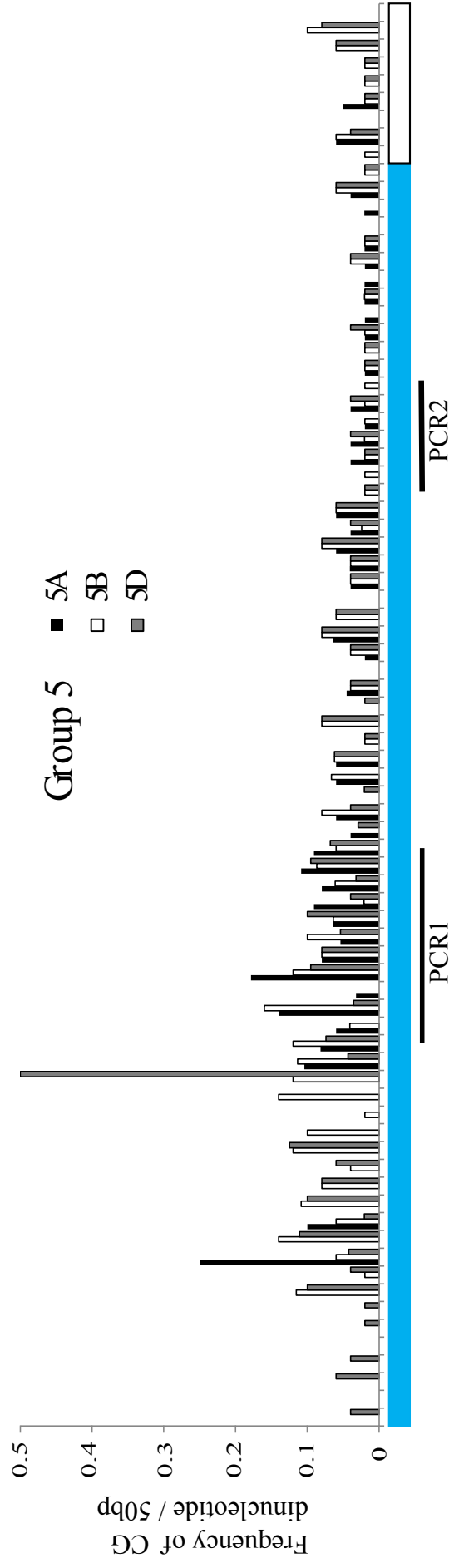
Additional file 2

	2B		5B		7A	
Position	ITMI	Position/top	ITMI	Position/top	ITMI	Position/top
proximal	Xgpw4382-2B	60.4	Xgwm604-5B	101.5	Xcfa2040-7A	226.5
distal	Xgpw2225-2B	60	Xfbb323-5B	121.2	Xgpw4050-7A	275.8

Additional file 3

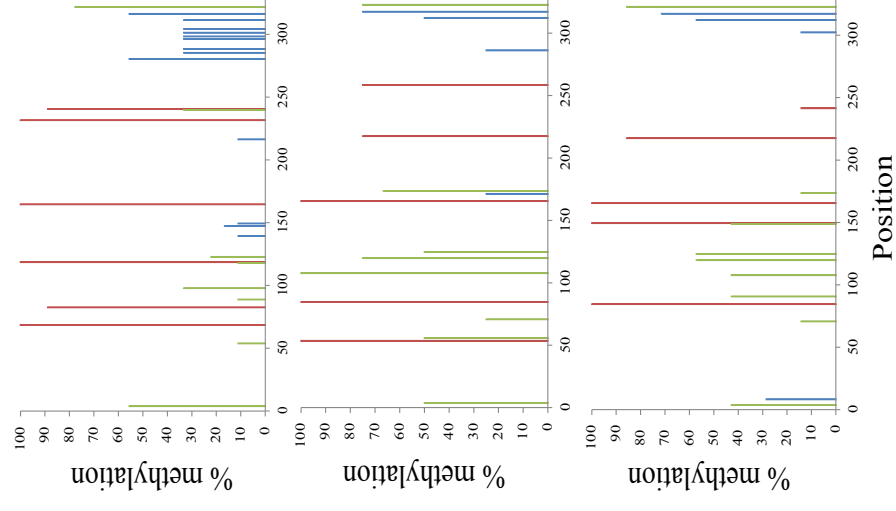


Additional file 4



CG
CHG
CHH

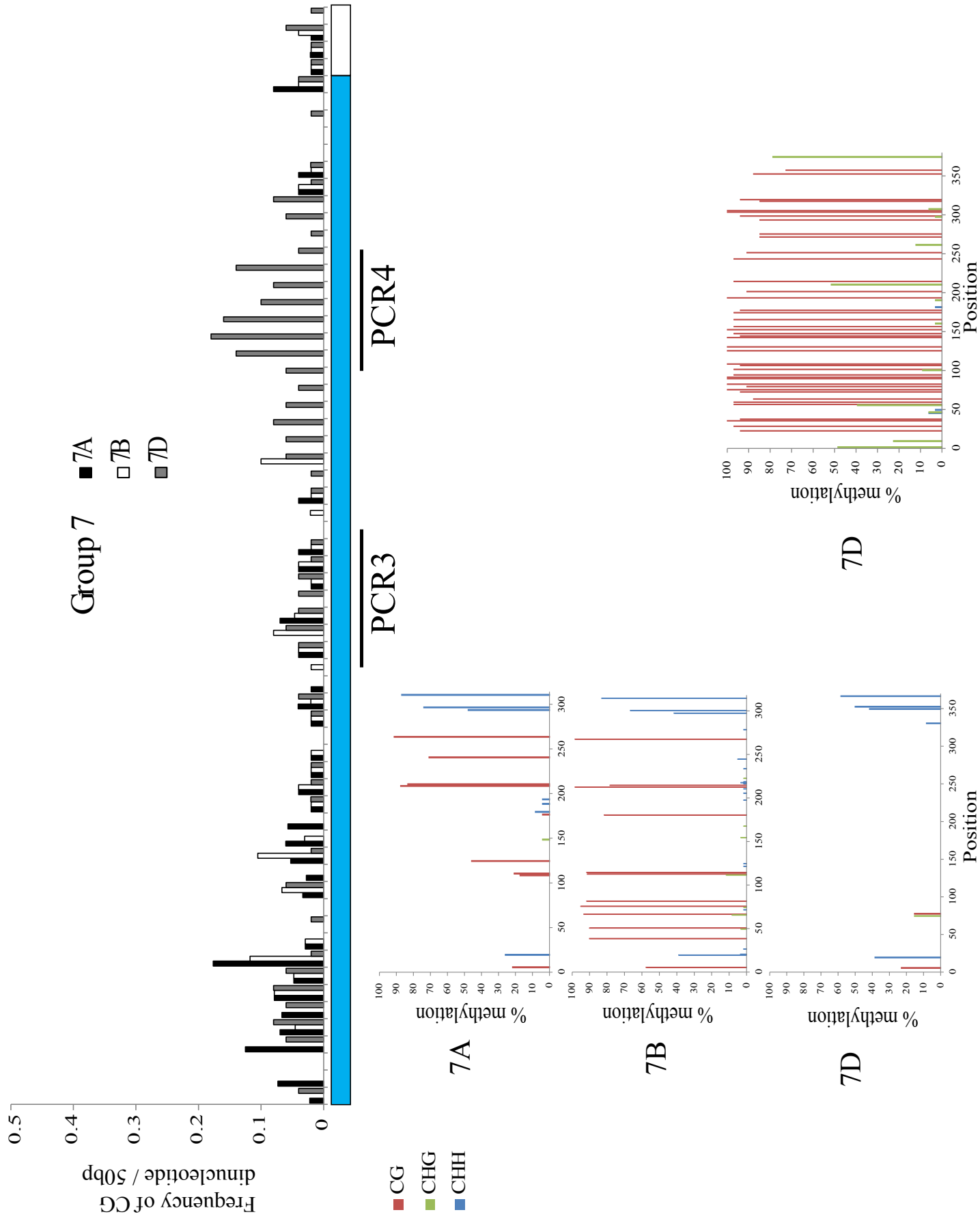
5A



5B

5D

Additional file 5



Additional file 6

Gene	Homoeolog specificity	Methods	Forward primer (5' to 3')	reverse primer (5' to 3')	Remarks
MET1-group2	2BD	BS	TGATTAYTTGGAGTGGIATGG	TTTCACAACRTRTTTTCTCTTCC	
	2A	BS	GYTTTAYTTGGAGTGGTATGAA	TTTCACAACRTRTTTTCTCTTCC	
	2A	RT-PCR	GATAGGATTATCACAGTCCGC	GCAAGAGGGGGTGGCACA	
	2B	RT-PCR	TGTGCCACCCCTCTTGC	CAGTTCATGCCAGCCATGC	
	2D	RT-PCR	GACCAGGATAGGATTATCAGC	CCAGCCACGCACAATAAAGA	
	2A	Q-RT-PCR	CAGTTTTCGGGCACCATTC	CCAGTCACGCACAATAGAGG	
	2B	Q-RT-PCR	CAGTTTTCGGGCACCATTC	CAGTTCATGCCAGCCATGC	
	2D	Q-RT-PCR	TGTGCCACCCCTCTTGC	CCAGCCACGCACAATAAAGA	
	2B	Mapping	GAAGGTGACCAAGTTCATGCTCA GTTTCAGGGTAGCAAACACG	GCTCCTGGAGAGATTCTTCTCTGATT	marker 52, Population Renan x Recital, SNP =G
2B	Mapping	GAAGTTCGGAGTCAACGGATTTC AGTTCAGGGTAGCAAACACT	TTCTTCTGATTGGCCAGAACCAT	marker 52, Population Renan x Recital, SNP =T	
MET1-group5	5A	Mapping	GAAGGTGACCAAGTTCATGCTCA CTTGTGCTCGGCAACTACAA	GGGAACAAAATCTACTGTCAACTATG CTT	marker 51, Population Renan x Recital, SNP =A
	5A	Mapping	GAAGTTCGGAGTCAACGGATTAC TTGTGCTCGGCAACTACAC	CTACTGTCAACTATGCTTCATTTTAG CCAA	marker 51, Population Renan x Recital, SNP =C
	5B	Mapping	GAAGTTCGGAGTCAACGGATTAC ATTCTTAGTAAGCAACCATGTGT	CCTGTGGATAGCTTACCTGTACTA	marker 35, Population Renan x Recital, SNP =A
	5B	Mapping	GAAGTTCGGAGTCAACGGATTTC CAGATGAACGCCCTTTTCCTA	ATTCCGCATTCTAGAAGCAGGTTTCA T	marker 43, Population Renan x Recital, SNP =A
	5B	Mapping	GAAGTTCGGAGTCAACGGATTTC GGCGAGCCCATCTCCTG	GCAAAAAGCGTATCGCCTCGAATCTT	marker 48, Population Renan x Recital, SNP =C
	5B	Mapping	GAAGTTCGGAGTCAACGGATTTC ATTCTTAGTAAGCAACCATGTGC	GCTTACCTGTACTATTACATGATG AAA	marker 35, Population Renan x Recital, SNP =G
	5B	Mapping	GAAGTTCGGAGTCAACGGATTTC AGATGAACGCCCTTTTCCTG	CAGGTTTCATATGGCATCGCGCATT	marker 43, Population Renan x Recital, SNP =G
	5B	Mapping	GAAGTTCGGAGTCAACGGATTTC AGCGGAGCCCATCTCCTA	GTATCGCCTCGAATCTTCTTCTGGAA	marker 48, Population Renan x Recital, SNP =T
	5B	BS	TTYATGYAAGGAAGTGGYAG	AACTCCTTTACCTTTTTRTCATC	
	5B	BS	GTTTTGYGGGAGGATTG	TCRRCCATCATTCATCC	
	5B	BS	GGTGGGTTYAAYTTTTGAT	AACCACATRTCCATRTAAAACC	
	5D	BS	TTYATGYAAGGAAGTGGYGG	TCCTTCCCTTTTTRTTRTCAC	
	5D	BS	GAAYTGTGAAGGAYTATYTG	ARATCTCRACAAAACACTCAT	
	5ABD	BS	ATYTGAAAYTTYTAATTAGGATGT	TCCCCTRTAAARTTTARAATCA	
	5A	RT-PCR	TCGATGATCTGCATATCTATTTC	ACTGCTTCGTTGGCTGGAG	
	5B	RT-PCR	CAGGAAGATCAAGGGATTCC	ACTGCTTCGTTGGCTGGAG	
5D	RT-PCR	GTGCTAAAAATGATACTGTTATC	ACTGCTTCGTTGGCTGGAG		
7A	7A	Mapping	GAAGGTGACCAAGTTCATGCTAT GCATCTTCGCTGAGTGATCAC	TGAGATTGAGCTCATTATTCTTTTA GCTA	marker 20, population Brigadier x Alcedo, SNP =A
	7A	Mapping	GAAGGTGACCAAGTTCATGCTAA CTCCACCGTTTGCCTGTGC	CAATCATGTGATGTACTGGYAGCTGA T	marker 22, Population Renan x Recital, SNP =G
	7A	Mapping	GAAGGTGACCAAGTTCATGCTAG CACCTGCCTCTAAAATCCAAAT	GCATTCCCTCATTCCACATGTTCTA AA	marker 24, Population Renan x Recital, SNP =A
	7A	Mapping	GAAGGTGACCAAGTTCATGCTGA TTGGCCTGAACCGATGCAC	GCATAGTATTTCCCTTCTGGTAGATT GAT	marker 25, Population Renan x Recital, SNP =C
	7A	Mapping	GAAGGTGACCAAGTTCATGCTAG TCAGGGATACAATTGGTGATCTC	TACTATACACCTACCGGTATTGTGG TIT	marker 26, Population Renan x Recital, SNP =C
	7A	Mapping	GAAGGTGACCAAGTTCATGCTCA ACAGAAGTCTGATTATGAAGT T	TATGGCCCTTCTGACTGAAAAGATCA AAA	marker 30, Population Renan x Recital, SNP =A
	7A	Mapping	GAAGGTGACCAAGTTCATGCTGG CGTAATTTGACCTGAAATC	CGATGAAAGCTCCCTCTAACTTACAT AT	marker 34, population Brigadier x Alcedo, SNP =G

MET1-group7	7A	Mapping	GAAGGTGACCAAGTTCATGCTGA TGCAATCATCAACATCACCACAC	TGACCATTGCGCTTTGCAGGGCAAT	marker 41, population Brigadier x Alcedo, SNP =C
	7A	Mapping	GAAGGTGGAGTCAACGGATTGC ATCTTCGCTGAGTGATCAGC	GCACCTTATGAGATTGAGCTCATICA TTT	marker 20, population Brigadier x Alcedo, SNP =G
	7A	Mapping	GAAGGTGGAGTCAACGGATTCA ACTCCACCGTTTGCCCTGTA	GCTGTGCGTGCAGGTGAAGCTA	marker 22, Population Renan x Recital, SNP =T
	7A	Mapping	GAAGGTGGAGTCAACGGATTCA CCTGCCTCTAAAATCCCAAC	TCCCACATGTTCTAAACTTCTCTGCA TTT	marker 24, Population Renan x Recital, SNP =G
	7A	Mapping	GAAGGTGGAGTCAACGGATTCT GATTGGCCTGAACCGATGCAT	TGATTTTCAGCTCAGGGCTGGAGAA	marker 25, Population Renan x Recital, SNP =T
	7A	Mapping	GAAGGTGGAGTCAACGGATTAG TCAGGATACAATTGGTGATCTG	GGTATTGTTGGTTGTGCGCACCAT	marker 26, Population Renan x Recital, SNP =G
	7A	Mapping	GAAGGTGGAGTCAACGGATTCA ACAGAACTAGCTGATTATGAATG C	AAGACCGGTAACACTTGTGGCTT	marker 30, Population Renan x Recital, SNP =G
	7A	Mapping	GAAGGTGGAGTCAACGGATTGG CTGGCGTAATTTGTACCTGAAAT A	CTCCCTCTAACTTACATATGGATCTC AAT	marker 34, population Brigadier x Alcedo, SNP =T
	7A	Mapping	GAAGGTGGAGTCAACGGATTGA TGCAATCATCAACATCACCACAG	GCCTTTGCAGGGCAATAATGGACAA	marker 41, population Brigadier x Alcedo, SNP =G
	7B	Mapping	GAAGGTGACCAAGTTCATGCTGC ATTTATTTCAACAATGCAGCGC	TGAAGAATTATCTGTCCATGCACCAC AA	marker 04, population Brigadier x Alcedo, SNP =C
	7B	Mapping	GAAGGTGACCAAGTTCATGCTAT GCACCACAACATATCGTTCAACG	AATTTGAGAAAGGGCAGCCTGCATTT ATT	marker 05, population Brigadier x Alcedo, SNP =C
	7B	Mapping	GAAGGTGACCAAGTTCATGCTCA CTTGCTTCGAATTTCAAACAGTC ATT	CTGTCTTGGCAATGCTGCAAAATGAA	marker 06, population Brigadier x Alcedo, SNP =A
	7B	Mapping	GAAGGTGACCAAGTTCATGCTTG CTCTCCTGTTTGTCCCTA	AAAAATGCAGAGGTTTAGAACATGT GGAA	marker 09, population Brigadier x Alcedo, SNP =A
	7B	Mapping	GAAGGTGACCAAGTTCATGCTGA GAGGGAGCGCGCACG	ATTTGCTGGCAACATCCAGAGCAA	marker 11, population Brigadier x Alcedo, SNP =C
	7B	Mapping	GAAGGTGGAGTCAACGGATTGC ATTTATTTCAACAATGCAGCGG	CTGTCCATGCACCACAACATATCGTT	marker 04, population Brigadier x Alcedo, SNP =G
	7B	Mapping	GAAGGTGGAGTCAACGGATTCA TGCACCACAACATATCGTTCAAC A	AAGGGCAGCCTGCATTTATTTCAACA AAT	marker 05, population Brigadier x Alcedo, SNP =T
	7B	Mapping	GAAGGTGGAGTCAACGGATTAC TTGCTTCGAATTTCAAACAGTCA TC	GCAATGCTGCAAAATGAACGAGAGGTT	marker 06, population Brigadier x Alcedo, SNP =G
	7B	Mapping	GAAGGTGGAGTCAACGGATTGC TCTCCTGTTTGTCCCTG	GAGGTTTAGAACATGTGGAAATGAAG GAA	marker 09, population Brigadier x Alcedo, SNP =G
	7B	Mapping	GAAGGTGGAGTCAACGGATTGA GAGGGAGCGCGCAC	TCCAGAGCAAGCACCAGGAGAT	marker 11, population Brigadier x Alcedo, SNP =G
	7A	BS	GGGTYGAGTTTGTATAAYTTA	AAARCCAAAACAATAARAARAC	
	7B	BS	GGGTYGAGTTTGTATAAYYTA	AAARCCAAAACAATAARAARAC	
	7D	BS	GGGTYGAGTTTGTATAAYATA	AAARCCAAAACAATAARAARAC	
	7D	BS	YGGGTGGYAGAGAAGAATG	ACACARACTCCTCTAAATTCAC	
	7A	RT-PCR	CATGCTTGTGGAGAACTG	ATCTTTGCAAGAAGTCCAGTG	
	7A	RT-PCR	CAGGATCATCACTGTTTCGTG	AATGACGAGCCTTGTATCATG	
	7D	RT-PCR	CAGGATCATCACTGTTTCGTG	GGCGACCAGCCTGTATATAG	
	7D	RT-PCR	GCTCAAGCAAGCCATCGAC	GGCGACCAGCCTGTATATAG	
VRN-A1	A	BS	AAATGATTTGGGAAAGCAAAT C	GAGGGGTTTTAAAGGGATCTGGCCC	0.0 k fragment from Khan et al 2013
	A	BS	GGAYAAAATYTTGAAYGGTATG AG	TTRCTAARCCCTTCAAAAACCTA	9.2 k fragment from Khan et al 2013

Additional file 7

CHROMOSOME ARM	IWGC CONTIG	IWGC ORDER	Os NAME	Os ORDER	Bd NAME	Bd ORDER
5AL	5AL:2744104	1	Os03g58040.1	1	Bradi1g05680.1	1
5AL	5AL:2732873	2	Os03g58050.1	2	Bradi1g05670.1	2
5AL	5AL:2734524	3	Os03g58060.1	3	Bradi1g05660.1	3
5AL	5AL:2732731	4	Os03g58070.1	4	Bradi1g05650.1	4
5AL	5AL:2736066	5	Os03g58080.1	5	Bradi1g05630.1	5
5AL	5AL:2695737	6	Os03g58090.1	6	Bradi1g05620.1	6
5AL	5AL:2763495	7	Os03g58100.1	7	Bradi1g05610.1	7
5AL	5AL:343725	8	Os03g58110.1	8	Bradi1g05600.1	8
5AL	5AL:2686684	9	Os03g58120.1	9	Bradi1g05590.1	9
5AL	5AL:353991	10	Os03g58130.1	10	Bradi1g05580.1	10
5AL	5AL:2760754	11	Os03g58140.1	11	Bradi1g05570.1	11
5AL	5AL:2760754	12	Os03g58150.1	12	Bradi1g05560.1	12
5AL	5AL:683544	13	Os03g58160.1	13	Bradi1g05550.1	13
5AL	5AL:2800470	14	Os03g58170.1	14		
5AL	5AL:2102481	15			Bradi1g05540.1	14
5AL	5AL:2694483	16	Os03g58190.1	15	Bradi1g05520.1	15
5AL	5AL:2691246	17	Os03g58204.1	16	Bradi1g05510.1	16
5AL	5AL:1744178	18	Os03g58240.1	17	Bradi1g05490.1	17
5AL	5AL:2741360	19	Os03g58250.1	18	Bradi1g05480.1	18
5AL	5AL:2801238	20	Os03g58260.1	19		
5AL	5AL:2759119	21			Bradi1g05470.1	19
5AL	5AL:1126042	22	Os03g58290.1	20	Bradi1g05460.1	20
5AL	5AL:1126042	23	Os03g58300.1	21	Bradi1g05450.1	21
5AL	5AL:1126042	24	Os03g58320.1	22		
5AL	5AL:2796056	25	Os03g58330.1	23	Bradi1g05430.1	22
5AL	5AL:2802820	26	Os03g58340.1	24	Bradi1g05420.1	23
5AL	5AL:2802820	27			Bradi1g05410.1	24
5AL	5AL:2799595	28	Os03g58390.1	25	Bradi1g05390.1	25
5AL	5AL:2806153	29	Os03g58400.1	26		
5AL	5AL:2672180	30			Bradi1g05380.1	26
5AL	5AL:32991	31	Os03g58420.1	27	Bradi1g05350.1	27
5AL	5AL:2734204	32	Os03g58430.1	28	Bradi1g05340.1	28
5AL	5AL:1181641	33	Os03g58480.1	29	Bradi1g05320.1	29
5AL	5AL:361545	34	Os03g58520.1	30	Bradi1g05260.1	30
5AL	5AL:429134	35	Os03g58530.1	31	Bradi1g05250.1	31
5AL	5AL:2376817	36	Os03g58540.1	32	Bradi1g05240.1	32
5AL	5AL:2741481	37			Bradi1g05200.1	33
5AL	5AL:1594861	38	Os03g58570.1	33	Bradi1g05190.1	34
5AL	5AL:2789791	39	Os03g58590.1	34	Bradi1g05170.1	35
5AL	5AL:2748163	40	Os03g58600.1	35	Bradi1g05160.1	36
5AL	5AL:2352467	41	Os03g58620.1	36		
5AL	5AL:2809517	42	Os03g58630.1	37		
5AL	5AL:2803096	43			Bradi1g05150.1	37
5AL	5AL:2798135	44			Bradi1g05140.1	38
5AL	5AL:2806991	45			Bradi1g05130.1	39
5AL	5AL:941218	46			Bradi1g05120.1	40
5AL	5AL:2768199	47			Bradi1g05110.1	41

TaMet-5A1

TaMet-5A1

5AL	5AL:2806991	48	Os03g58640.1	38		
5AL	5AL:598579	49	Os03g58670.1	39		
5AL	5AL:2782620	50	Os03g58700.1	40		
5AL	5AL:2809814	51	Os03g58710.1	41		
5AL	5AL:2768199	52	Os03g58720.1	42		
5AL	5AL:828164	53	Os03g58740.1	43	Bradi1g05100.1	42
5AL	5AL:2353167	54			Bradi1g05080.1	43
5AL	5AL:1685317	55	Os03g58780.1	44	Bradi1g05070.1	44
5AL	5AL:2805920	56	Os03g58820.1	45	Bradi1g05060.1	45
5AL	5AL:2796269	57	Os03g58830.1	46	Bradi1g05040.1	46
5AL	5AL:1230154	58	Os03g58840.1	47	Bradi1g05030.1	47
5AL	5AL:2741244	59	Os03g58850.1	48		
5AL	5AL:2809741	60			Bradi1g05020.1	48
5AL	5AL:2682679	61	Os03g58870.1	49	Bradi1g05010.1	49
5AL	5AL:986955	62	Os03g58890.1	50	Bradi1g05000.1	50
5AL	5AL:1907283	63	Os03g58900.1	51	Bradi1g04990.1	51
5AL	5AL:2783401	64	Os03g58910.2	52	Bradi1g04980.1	52
5AL	5AL:2805442	65			Bradi1g04960.1	53
5BL	5BL:10816777	1	Os03g58040.1	1	Bradi1g05680.1	1
5BL	5BL:10797603	2	Os03g58050.1	2	Bradi1g05670.1	2
5BL	5BL:10806818	3	Os03g58060.1	3	Bradi1g05660.1	3
5BL	5BL:10721087	4	Os03g58070.1	4	Bradi1g05650.1	4
5BL	5BL:10824935	5			Bradi1g05640.1	5
5BL	5BL:8729501	6	Os03g58080.1	5	Bradi1g05630.1	6
5BL	5BL:2853232	7	Os03g58090.1	6	Bradi1g05620.1	7
5BL	5BL:10806811	8	Os03g58100.1	7	Bradi1g05610.1	8
5BL	5BL:10806811	9	Os03g58110.1	8	Bradi1g05600.1	9
5BL	5BL:10887964	10	Os03g58120.1	9	Bradi1g05590.1	10
5BL	5BL:10845579	11	Os03g58130.1	10	Bradi1g05580.1	11
5BL	5BL:10845579	12	Os03g58140.1	11	Bradi1g05570.1	12
5BL	5BL:10845579	13	Os03g58150.1	12	Bradi1g05560.1	13
5BL	5BL:10845579	14	Os03g58160.1	13	Bradi1g05550.1	14
5BL	5BL:10920668	15	Os03g58170.1	14	Bradi1g05540.1	15
5BL	5BL:10890955	16	Os03g58180.1	15	Bradi1g05530.1	16
5BL	5BL:10899556	17	Os03g58190.1	16	Bradi1g05520.1	17
5BL	5BL:10837461	18	Os03g58204.1	17	Bradi1g05510.1	18
5BL	5BL:10826094	19	Os03g58230.1	18	Bradi1g05500.1	19
5BL	5BL:10792150	20	Os03g58240.1	19	Bradi1g05490.1	20
5BL	5BL:10189661	21	Os03g58250.1	20	Bradi1g05480.1	21
5BL	5BL:10839350	22	Os03g58260.1	21	Bradi1g05470.1	22
5BL	5BL:10899718	23	Os03g58290.1	22	Bradi1g05460.1	23
5BL	5BL:10899718	24			Bradi1g05450.1	24
5BL	5BL:10807820	25	Os03g58300.1	23		
5BL	5BL:10924991	26	Os03g58320.1	24		
5BL	5BL:10825701	27	Os03g58330.1	25		
5BL	5BL:10825701	28			Bradi1g05430.1	25
5BL	5BL:10921140	29	Os03g58340.1	26	Bradi1g05420.1	26
5BL	5BL:10921140	30	Os03g58350.1	27	Bradi1g05410.1	27
5BL	5BL:10864426	31	Os03g58380.1	28		
5BL	5BL:7717076	32	Os03g58390.1	29	Bradi1g05390.1	28
5BL	5BL:10925378	33	Os03g58400.1	30	Bradi1g05380.1	29
5BL	5BL:10924332	34	Os03g58410.1	31	Bradi1g05370.1	30

TaMet-5B1

5BL	5BL:10867378	35	Os03g58420.1	32	Bradi1g05350.1	31
5BL	5BL:10795568	36	Os03g58430.1	33	Bradi1g05340.1	32
5BL	5BL:10822114	37			Bradi1g05330.1	33
5BL	5BL:10825924	38	Os03g58480.1	34	Bradi1g05320.1	34
5BL	5BL:10825924	39			Bradi1g05310.1	35
5BL	5BL:10893769	40	Os03g58500.1	35		
5BL	5BL:10814367	41			Bradi1g05280.1	36
5BL	5BL:10879097	42	Os03g58520.1	36	Bradi1g05260.1	37
5BL	5BL:10879097	43	Os03g58530.1	37	Bradi1g05250.1	38
5BL	5BL:10838104	44	Os03g58540.1	38	Bradi1g05240.1	39
5BL	5BL:10838104	45			Bradi1g05200.1	40
5BL	5BL:10840323	46	Os03g58570.1	39	Bradi1g05190.1	41
5BL	5BL:10844621	47	Os03g58580.1	40	Bradi1g05180.1	42
5BL	5BL:10914845	48	Os03g58590.1	41	Bradi1g05170.1	43
5BL	5BL:10914845	49			Bradi1g05160.1	44
5BL	5BL:10816505	50	Os03g58600.1	42		
5BL	5BL:10829640	51	Os03g58620.1	43		
5BL	5BL:10922830	52	Os03g58630.1	44	Bradi1g05150.1	45
5BL	5BL:10832199	53	Os03g58670.1	46		
5BL	5BL:10801521	54	Os03g58710.1	48	Bradi1g05140.1	46
5BL	5BL:10882913	55	Os03g58640.1	45	Bradi1g05130.1	47
5BL	5BL:10815386	56	Os03g58700.1	47	Bradi1g05120.1	48
5BL	5BL:10879461	57	Os03g58720.1	49	Bradi1g05110.1	49
5BL	5BL:10855303	58	Os03g58740.1	50	Bradi1g05100.1	50
5BL	5BL:10918676	59	Os03g58750.1	51	Bradi1g05090.1	51
5BL	5BL:10865573	60			Bradi1g05080.1	52
5BL	5BL:10789814	61	Os03g58780.1	52	Bradi1g05070.1	53
5BL	5BL:10735609	62	Os03g58790.1	53		
5BL	5BL:10735609	63	Os03g58800.1	54		
5BL	5BL:10787074	64	Os03g58810.1	55		
5BL	5BL:10840449	65	Os03g58820.1	56	Bradi1g05060.1	54
5BL	5BL:1283675	66	Os03g58830.1	57	Bradi1g05040.1	55
5BL	5BL:10867098	67	Os03g58840.1	58	Bradi1g05030.1	56
5BL	5BL:10824553	68	Os03g58850.1	59		
5BL	5BL:10893855	69			Bradi1g05020.1	57
5BL	5BL:10905571	70	Os03g58870.1	60	Bradi1g05010.1	58
5BL	5BL:10866215	71	Os03g58890.1	61	Bradi1g05000.1	59
5BL	5BL:10871750	72	Os03g58900.1	62	Bradi1g04990.1	60
5BL	5BL:10923312	73	Os03g58910.2	63	Bradi1g04980.1	61
5BL	5BL:10871750	74	Os03g58920.1	64	Bradi1g04960.1	62
5DL	5DL:4579940	1	Os03g58040.1	1	Bradi1g05680.1	1
5DL	5DL:4568611	2	Os03g58050.1	2	Bradi1g05670.1	2
5DL	5DL:4552307	3	Os03g58060.1	3	Bradi1g05660.1	3
5DL	5DL:4503194	4			Bradi1g05640.1	4
5DL	5DL:4532486	5	Os03g58080.1	4	Bradi1g05630.1	5
5DL	5DL:4532486	6	Os03g58090.1	5	Bradi1g05620.1	6
5DL	5DL:4603458	7	Os03g58100.1	6	Bradi1g05610.1	7
5DL	5DL:4603458	8	Os03g58110.1	7	Bradi1g05600.1	8
5DL	5DL:4541028	9	Os03g58120.1	8	Bradi1g05590.1	9
5DL	5DL:4513175	10	Os03g58130.1	9	Bradi1g05580.1	10
5DL	5DL:4513175	11	Os03g58140.1	10	Bradi1g05570.1	11
5DL	5DL:4513175	12	Os03g58150.1	11	Bradi1g05560.1	12

5DL	5DL:4513175	13	Os03g58160.1	12	Bradi1g05550.1	13
5DL	5DL:4524567	14	Os03g58170.1	13	Bradi1g05540.1	14
5DL	5DL:4500384	15	Os03g58190.1	14	Bradi1g05520.1	15
5DL	5DL:857395	16	Os03g58204.1	15	Bradi1g05510.1	16
5DL	5DL:4592014	17	Os03g58230.1	16	Bradi1g05500.1	17
5DL	5DL:4603599	18	Os03g58240.1	17	Bradi1g05490.1	18
5DL	5DL:4603599	19	Os03g58250.1	18	Bradi1g05480.1	19
5DL	5DL:4600688	20	Os03g58260.1	19	Bradi1g05470.1	20
5DL	5DL:4547569	21	Os03g58300.1	20		
5DL	5DL:4547569	22	Os03g58320.1	22	Bradi1g05460.1	21
5DL	5DL:4496471	23	Os03g58290.1	23	Bradi1g05450.1	22
5DL	5DL:4515833	24	Os03g58330.1	21	Bradi1g05430.1	23
5DL	5DL:4490001	25	Os03g58340.1	24	Bradi1g05420.1	24
5DL	5DL:4490001	26	Os03g58350.1	25	Bradi1g05410.1	25
5DL	5DL:3393307	27	Os03g58390.1	26	Bradi1g05390.1	26
5DL	5DL:4608210	28	Os03g58400.1	27	Bradi1g05380.1	27
5DL	5DL:4567022	29	Os03g58410.1	28	Bradi1g05370.1	28
5DL	5DL:4566006	30	Os03g58420.1	29	Bradi1g05350.1	29
5DL	5DL:4524187	31	Os03g58430.1	30	Bradi1g05340.1	30
5DL	5DL:4556738	32			Bradi1g05330.1	31
5DL	5DL:2336334	33	Os03g58480.1	33	Bradi1g05320.1	32
5DL	5DL:2910741	34	Os03g58500.1	34		
5DL	5DL:4571477	35			Bradi1g05290.1	33
5DL	5DL:4571477	36			Bradi1g05270.1	34
5DL	5DL:1047956	37			Bradi1g05280.1	35
5DL	5DL:4608660	38	Os03g58520.1	35	Bradi1g05260.1	36
5DL	5DL:4473836	39	Os03g58530.1	36	Bradi1g05250.1	37
5DL	5DL:4495237	40	Os03g58540.1	37	Bradi1g05240.1	38
5DL	5DL:4495237	41			Bradi1g05200.1	39
5DL	5DL:4572451	42	Os03g58570.1	38	Bradi1g05190.1	40
5DL	5DL:4532064	43	Os03g58580.1	39	Bradi1g05180.1	41
5DL	5DL:4532779	44	Os03g58590.1	40	Bradi1g05170.1	42
5DL	5DL:4557608	45	Os03g58600.1	41	Bradi1g05160.1	43
5DL	5DL:4546142	46	Os03g58620.1	42		
5DL	5DL:4571419	47	Os03g58630.1	43	Bradi1g05150.1	44
5DL	5DL:4571419	48	Os03g58640.1	44	Bradi1g05140.1	45
5DL	5DL:4571419	49			Bradi1g05130.1	46
5DL	5DL:4523323	50	Os03g58700.1	45	Bradi1g05120.1	47
5DL	5DL:4608197	51	Os03g58710.1	46		
5DL	5DL:4541544	52	Os03g58720.1	47	Bradi1g05110.1	48
5DL	5DL:4517824	53	Os03g58740.1	48	Bradi1g05100.1	49
5DL	5DL:4603756	54	Os03g58780.1	49	Bradi1g05080.1	50
5DL	5DL:4489274	55	Os03g58790.1	50		
5DL	5DL:4517163	56	Os03g58800.1	51		
5DL	5DL:4585906	57	Os03g58810.1	52		
5DL	5DL:4509470	58	Os03g58820.1	53		
5DL	5DL:4603756	59			Bradi1g05070.1	51
5DL	5DL:4509470	60			Bradi1g05060.1	52
5DL	5DL:4606041	61			Bradi1g05050.1	53
5DL	5DL:4489431	62	Os03g58830.1	54	Bradi1g05040.1	54
5DL	5DL:4572971	63	Os03g58840.1	55	Bradi1g05030.1	55
5DL	5DL:4551533	64			Bradi1g05020.1	56

TaMet-5D1

5DL	5DL:4565592	65	Os03g58850.1	56		
5DL	5DL:4599310	66	Os03g58870.1	57		
5DL	5DL:4599310	67	Os03g58890.1	58		
5DL	5DL:4570311	68	Os03g58900.1	59		
5DL	5DL:4496706	69	Os03g58910.2	60		
5DL	5DL:4557991	70	Os03g58920.1	61		
7AL	7AL:4500764	1	Os03g58040.1	1	Bradi1g05680.1	1
7AL	7AL:4535568	2	Os03g58050.1	2	Bradi1g05670.1	2
7AL	7AL:4532056	3	Os03g58400.1	3	Bradi1g05380.1	3
7AL	7AL:4441897	4			Bradi1g05360.1	4
7AL	7AL:4534795	5	Os03g58430.1	5	Bradi1g05340.1	5
7AL	7AL:1206673	6			Bradi1g05210.1	6
7AL	7AL:4452921	7	Os03g58570.1	7	Bradi1g05190.1	7
7AL	7AL:4510505	8	Os03g58580.1	8	Bradi1g05180.1	8
7AL	7AL:4492195	9	Os03g58630.1	9	Bradi1g05150.1	9
7AL	7AL:4492195	10			Bradi1g05140.1	10
7AL	7AL:4483901	11	Os03g58670.1	11		
7AL	7AL:4472017	12	Os03g58790.1	12		
7AL	7AL:4472017	13	Os03g58800.1	13		
7AL	7AL:4472191	14	Os03g58810.1	14		
7AL	7AL:4472620	15	Os03g58830.1	15	Bradi1g05040.1	11
7BL	7BL:6701310	1	Os03g58040.1	1	Bradi1g05680.1	1
7BL	7BL:6694785	2	Os03g58050.1	2	Bradi1g05670.1	2
7BL	7BL:6665951	3	Os03g58160.1	3	Bradi1g05550.1	3
7BL	7BL:6682174	4	Os03g58400.1	4	Bradi1g05380.1	4
7BL	7BL:279551	5	Os03g58670.1	6		
7BL	7BL:6743224	6	Os03g58790.1	7		
7BL	7BL:6674971	7	Os03g58800.1	8		
7BL	7BL:6697122	8			Bradi1g05360.1	5
7BL	7BL:6501267	9			Bradi1g05350.1	6
7BL	7BL:6710382	10			Bradi1g05290.1	7
7BL	7BL:6632036	11			Bradi1g05280.1	8
7BL	7BL:6710382	12			Bradi1g05270.1	9
7BL	7BL:6745812	13			Bradi1g05210.1	10
7BL	7BL:6689829	14	Os03g58570.1	5	Bradi1g05190.1	11
7BL	7BL:6712874	15	Os03g58830.1	9	Bradi1g05040.1	12
7DL	7DL:3378869	1	Os03g58040.1	1	Bradi1g05680.1	1
7DL	7DL:3359841	2	Os03g58050.1	2	Bradi1g05670.1	2
7DL	7DL:3344608	3			Bradi1g05570.1	3
7DL	7DL:3392185	4	Os03g58400.1	3	Bradi1g05380.1	4
7DL	7DL:3392185	5	Os03g58410.1	4	Bradi1g05370.1	5
7DL	7DL:3374548	6			Bradi1g05360.1	6
7DL	7DL:2581233	7			Bradi1g05350.1	7
7DL	7DL:3368356	8			Bradi1g05210.1	8
7DL	7DL:3341999	9	Os03g58570.1	5	Bradi1g05190.1	9
7DL	7DL:3361750	10	Os03g58580.1	6	Bradi1g05180.1	10
7DL	7DL:3366313	11	Os03g58670.1	7		
7DL	7DL:3393412	12	Os03g58790.1	8		
7DL	7DL:3393412	13	Os03g58800.1	9		
7DL	7DL:3387834	14	Os03g58810.1	10		
7DL	7DL:3392066	15			Bradi1g05050.1	11

TaMet-7A1

TaMet-7B1

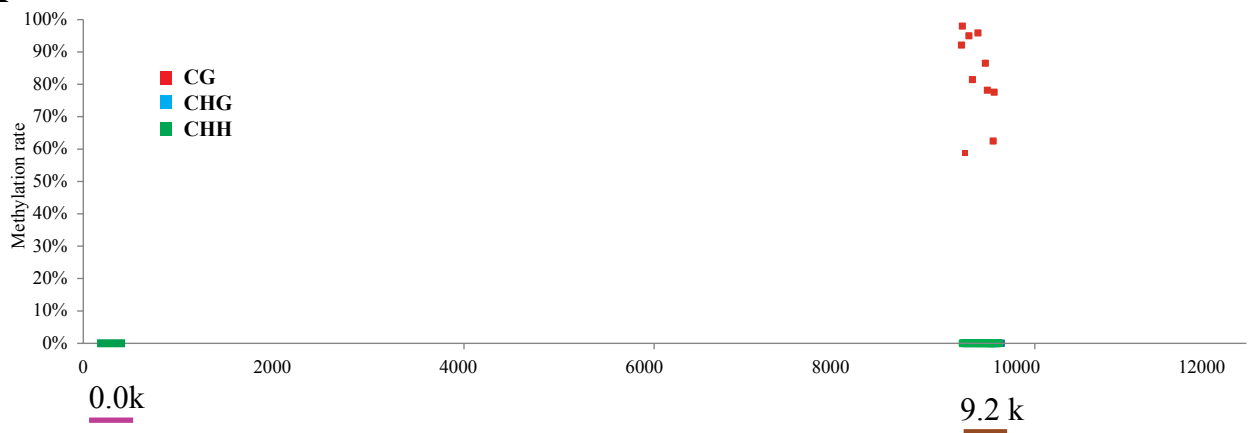
TaMet-7D1

Additional file 8

Stage	Wheat growth stage	Feekes scale	Zadoks scale	Leaves	Root	Stem	Spike	Grain
Seedling	First leaf through coleoptile	1	10	x	x			
Three leaves	3 leaves unfolded		13		x			
Three tillers	Main shoot and 3 tillers		23	x				
Spike at 1 cm	Pseudostem erection	5	30			x		
Two nodes	2nd detectable node	7	32			x	x	
Meiosis	Flag leaf ligule and collar visible	9	39		x		x	
Anthesis	1/2 of flowering complete		65			x	x	
2 DAAs (50°C.days)	Kernel (caryopsis) watery ripe		71	x				x
14 DAAs (350°C.days)	Medium Milk		75					x
30 DAAs (700°C.days)	Soft dough		85					x

Additional file 9

A



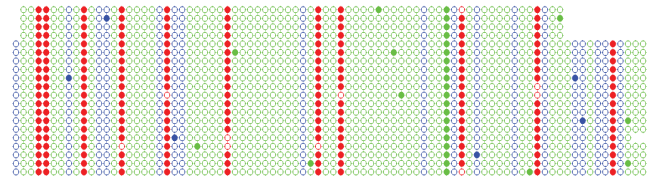
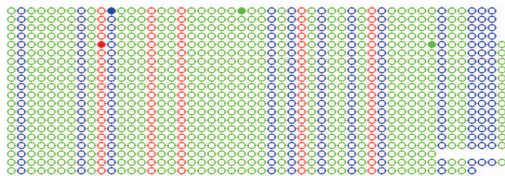
B



C

0.0 k (20 clones)

9.2 k (20 clones)



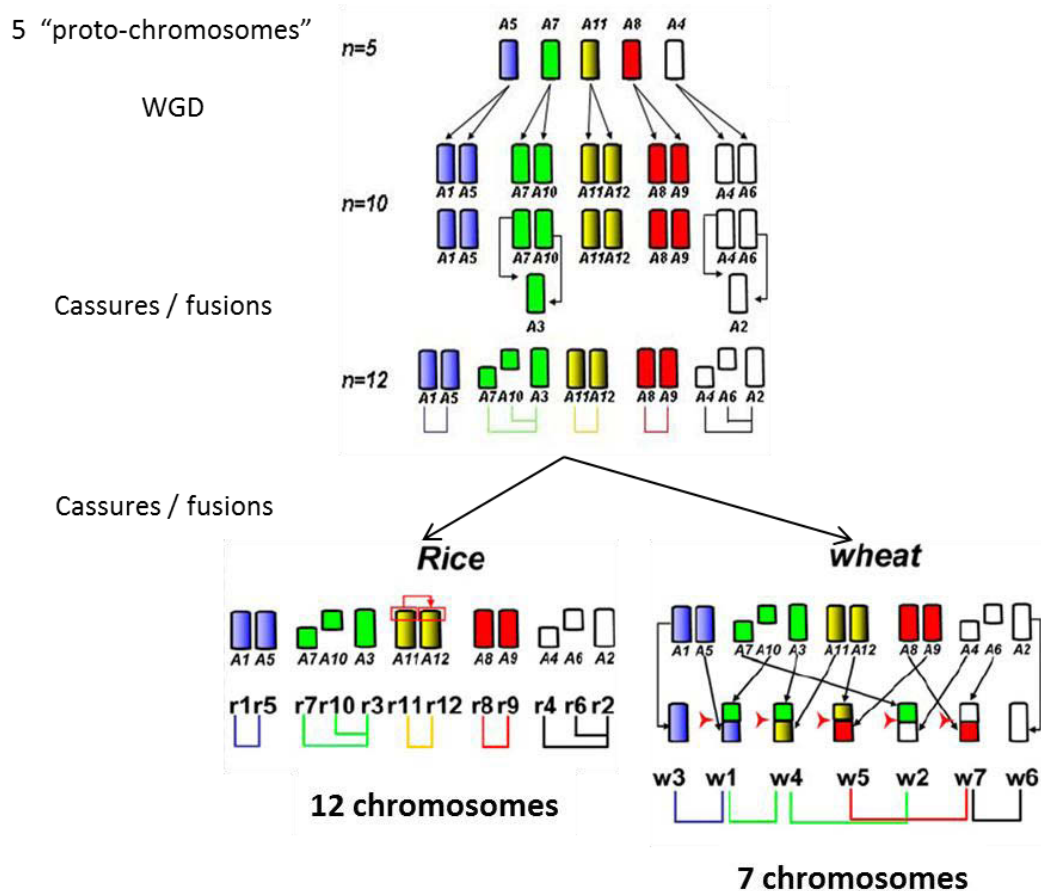


Figure 23: Histoire évolutive du génome du riz et du blé

Source : (Salse et al., 2008)

Les travaux récents de Paléogénomique chez les céréales ont établi les relations synténiques entre différentes graminées (riz, sorgho, Brachypodium et blé). D'après ces études, 7 régions synténiques seraient partagées entre le riz et le blé ce qui a permis d'émettre l'hypothèse d'un ancêtre commun à 5 chromosomes appelés A5, A7, A11, A8 et A4. Dans ce modèle, il y a environ 50-70MYa, l'ancêtre à n=5 « proto-chromosomes » aurait subi une duplication de l'ensemble du génome (Whole Genome Duplication; WGD) donnant lieu à une espèce intermédiaire avec n= 10. Des cassures/fusions seraient ensuite survenues générant des duplications de segments avec l'apparition d'une espèce intermédiaire à n= 12.

Cette situation a été bien conservée pour le riz encore segmenté en 12 chromosomes et dont r7 (qui porte *OsMET1a*), r3 (qui porte *OsMET1b*) et r10 sont issus du proto-chromosome A7. Chez le blé, il y aurait eu des cassures/fusions supplémentaires. Sur le schéma, il est encore possible d'identifier l'origine de w2 (qui porte l'orthologue de *OsMET1b*) qui provient également du proto-chromosome A7

5. Conclusion du chapitre 1

A partir des séquences des génomes du riz et de *Brachypodium*, ainsi que d'ESTs de blé nous avons défini la séquence des gènes *TaMET-1* en développant dans un premier temps la stratégie de Sequence Capture qui a ensuite été complétée par une analyse BLAST contre les séquences génomiques IWGSC. Le blé hexaploïde possède neuf copies de *TaMET-1* présentes sur les chromosomes 2, 5 et 7 et ce pour chacun des génomes homéologues A, B et D. La caractérisation de ces neuf gènes et des analyses de synténie avec les espèces proches permettent de conclure que tous n'ont pas la même fonction et la même histoire évolutive.

5.1. Histoire évolutive des gènes *MET1* chez les graminées

Le groupe 2, qui est le plus proche de *OsMET1b* chez le riz, semble correspondre aux copies fonctionnelles de *TaMET-1* chez le blé hexaploïde. Le profil d'expression de *MET1* est assez bien défini chez *Arabidopsis* et le riz où l'expression est forte dans les tissus reproducteurs (Annexes) selon les données Genevestigator. Chez le blé, les données RNA-Seq confirment une forte expression dans deux de nos tissus : d'une part la tige au stade Z30 qui, d'après nos discussions avec l'équipe d'Etienne Paux, inclus à la fois des tissus de la tige et ceux du méristème apical (il n'y a pas eu de dissection fine pour séparer tige et méristème) et d'autre part l'épi au stade Z32. Il y aurait donc aussi une expression dans les tissus reproducteurs chez le blé. Ceci reste toutefois à préciser plus finement dans le futur par des études en RT-PCR sur des tissus disséqués plus finement. Les relations de synténie entre la région génomique du blé portant *TaMET-1* des chromosome 2 (w2) et le riz montrent que w2 est synténique du chromosome 7 de riz (r7). Ce segment chromosomique serait également une région ancestrale portée par le proto-chromosome A7 (Salse et al., 2008) (Fig 22). Un élément très intéressant est le fait que le chromosome de riz r7 porte le gène *OsMET1b* (Os07g08500), le gène le plus exprimé du riz (Yamauchi et al., 2008b). Il y a donc une parfaite cohérence entre riz et blé : c'est le gène *MET1b* présent sur r7 et w2 qui est le plus exprimé. Ceci a également été confirmé dans nos toutes dernières analyses sur l'orge et sur les génomes diploïdes A (*T. urartu*) et D (*T. tauschii*).

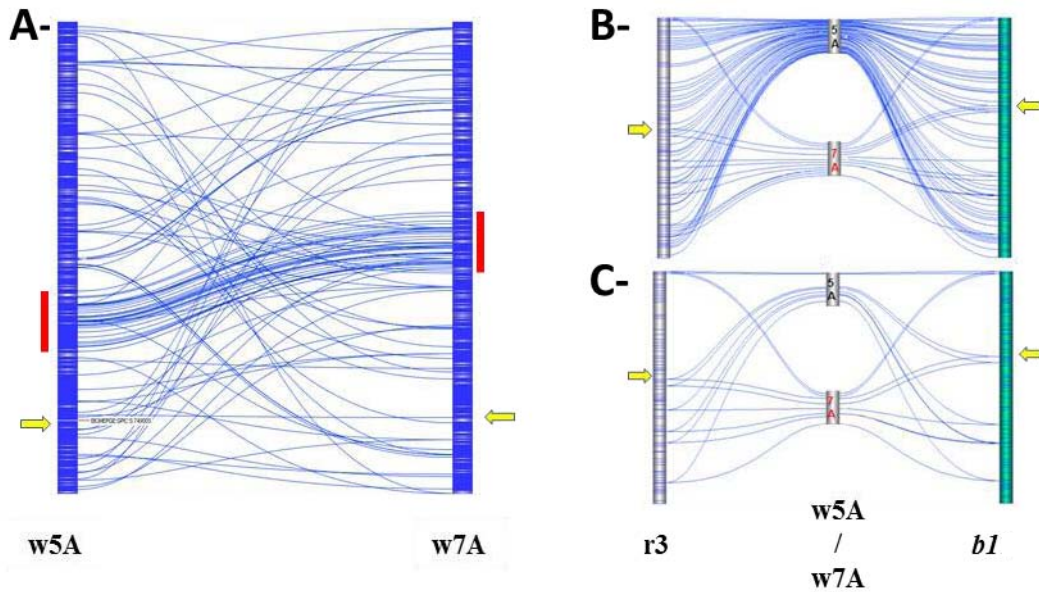


Figure 24 : Analyse de la synténie entre le riz, *Brachypodium* et le blé dans la région des copies *TaMET-1* des chromosomes 5 et 7

A- Une analyse BLASTn a été faite entre les ADNc prédits comme provenant des chromosomes 5A (w5A) et 7A (w7A) du blé (communications personnelles de Biogemma). Les fragments génomiques conservés entre les chromosomes 5A et 7A étaient déjà décrits dans Salse et *al.*, 2008. Cette région est facilement visualisable et indiquée par un trait rouge vertical, mais distincte de la région contenant *TaMET-1*.

B- Analyse par micro-synténie des régions génomiques des chromosomes 5A et 7A portant *TaMET1* avec les régions correspondantes du chromosomes 1 de *Brachypodium* (b1) et du chromosome 3 du riz (r3). Les flèches jaunes indiquent la présence des locus *MET1*.

C- Micro-synténie commune au chromosomes 5A et 7A du blé, le chromosomes 1 de *Brachypodium* et le chromosomes 3 du riz. Seulement 8 gènes sont conservés entre les deux chromosomes de blé.

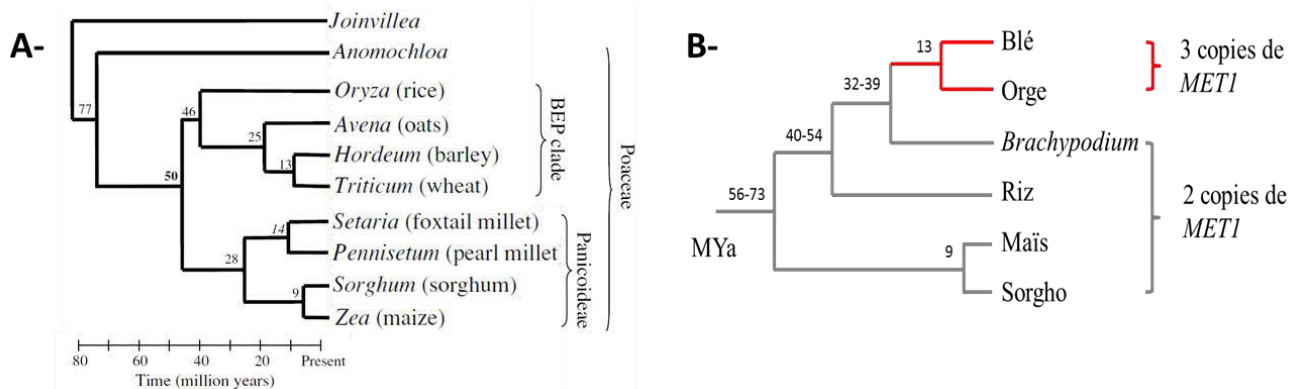


Figure 25: Datation des deux duplications ayant conduit à la formation de trois paralogues de *MET1* chez les *Triticeae*

A- Rappel sur l'origine des principales graminées selon Gaut, 2001

B- L'histoire évolutive des céréales montre qu'elles sont apparues il y a environ 56 à 73 Mya. Le premier événement de duplication a donc eu lieu après la spéciation des monocots, lors de la duplication de tout le génome (WGD) qui est commune à toutes ces espèces. La plupart des espèces comportent donc 2 copies de *MET1*. Seuls le blé et l'orge comportent une troisième copie. Le second événement de duplication a donc eu lieu un peu avant la divergence, donnant les copies des groupes 5 et 7.

D'un point de vue évolutif, l'évènement de duplication de l'ensemble du génome (Whole Genome Duplication, WGD) qui a eu lieu chez l'ancêtre des céréales a donné à partir de A7 tout d'abord A10 qui par cassure/fusion a donné les chromosomes A3 et A10 (Fig 23). Cette organisation à 12 chromosomes est bien conservée chez le riz : A10 correspond respectivement aux chromosomes 7 (r7), 3 (r3) et 10 (r10) de riz. En observant le riz, *OsMET1* se retrouve ainsi sur r7 et r3. Le paralogue de *MET1* appelé *OsMET1a* est la version la moins exprimée de *MET1* et se trouve sur le chromosome r3 (Annexes) (Yamauchi et al., 2008). A nouveau, en nous basant sur les travaux de l'équipe de Jérôme Salse, nous avons conclu que la région synténique de r3 se trouve sur le chromosome 5 de blé (w5). Les relations de synténie entre r3 et w5 sont d'ailleurs très fortes alors qu'elles sont très faibles avec w7 (Fig 24B). L'évènement de WGD commun à toutes les céréales explique donc la présence de deux copies chez le riz (r7 et r3), *Brachypodium* (bradi1a et bradi1b) et le blé (w2 et w5).

Enfin, un dernier évènement de duplication à l'échelle du gène ou d'un segment chromosomique court serait apparu plus récemment entre w5 et w7. Cette duplication est spécifique de la tribu des *Triticeae* (orge et blé) (Fig 25). D'un point de vue fonctionnel, les copies du groupe 7 sont certes fonctionnelles mais exprimées à des niveaux beaucoup plus faibles que celles du groupe 2. Les copies du groupe 5 sont quant à elles devenues des pseudogènes.

Ce travail a été compliqué à mettre en place car nous n'avions pas tous les outils d'analyses nécessaires ni les connaissances en évolution des génomes. Ce travail a été réalisé avec l'aide de Mickaël Throude (Biogemma), de Mickaël Abrouk et de Frédéric Choulet (INRA de Crouël). Cette partie de mon projet n'était pas non plus une priorité d'autant plus qu'il restait l'établissement des lignées RNAi (voir partie 2). Toutefois, ce travail a permis de compléter notre publication et donne une explication au fait que le génome du blé comporte neuf copies de notre gène d'étude.

5.2. Vitesse évolutive des gènes *MET1*

Une autre partie de notre analyse a été d'évaluer la vitesse évolutive de chaque copie de *MET1*. En général, après un évènement de duplication, la vitesse d'évolution des gènes est plus rapide puis tend à rediminuer. Étonnement, pour les gènes *TaMET-1*, ce ne sont pas les copies du groupe 7, qui sont apparues le plus récemment, qui ont évoluées le plus vite mais

celles du groupe 5. Les gènes du groupe 5 ont tous une vitesse d'évolution élevée caractéristique des pseudogènes.

Comment s'explique cette forte vitesse d'évolution pour le groupe 5 ? Plusieurs hypothèses peuvent être formulées parmi lesquelles, la méthylation, le niveau d'expression des gènes (Yang and Gaut, 2011) et la position chromosomique (Yang and Gaut, 2011, Etienne Paux et Lise Pingault communication personnelle) pourraient jouer un rôle important. Concernant la relation méthylation / transcription / vitesse évolutive, des études ont été réalisées chez *Arabidopsis* et montrent que sur 14 variables testées pour leur effet potentiel sur le taux de substitution non synonyme (dN dans notre publication), le niveau de transcription et la spécificité tissulaire sont les deux principaux facteurs (Yang and Gaut, 2011). Ainsi, une faible transcription est souvent associée à une plus forte vitesse évolutive. Une possibilité serait que l'arrivée de la méthylation sur le promoteur des gènes du groupe 5 soit nécessaire pour diminuer le niveau de transcription ce qui aurait pour conséquence l'augmentation de la vitesse évolutive. L'accumulation des mutations non-synonymes (donc de mutations délétères) entraîneraient à terme la pseudogénisation. Une fois converti en pseudogène, le besoin en méthylation serait toutefois moins « nécessaire ». Il y aurait donc au niveau des pseudogènes une diminution graduelle du niveau de méthylation. Un tel mécanisme a été rapporté récemment pour le devenir de gènes dupliqués pour le génome humain (Keller and Yi, 2014). Dans notre cas, il sera difficile d'aller plus loin dans l'élucidation de ces mécanismes évolutifs.

Concernant la relation entre la vitesse évolutive et la position chromosomique, il existe des positions à forte ou faible vitesse d'évolution. Le mécanisme n'est pas bien connu mais pourrait être en relation avec l'efficacité de recombinaison d'une région. Il semblerait que les données IWGSC permettent de formuler une hypothèse selon laquelle les gènes les plus conservés seraient plutôt en position proximale par rapport au centromère (tel que les gènes du groupes 2) alors que les gènes avec un profil d'expression présentant des spécificités tissulaires seraient en position plus distal. Enfin, les analyses de Lise Pingault tendent à démontrer que les régions distales sont enrichies en pseudogènes. Ceci est cohérent avec la position des gènes des groupes 5 et 7 à des positions distales (voir tableau 1 de la publication). Dans le futur, il serait donc intéressant de comparer les vitesses d'évolution des gènes adjacents aux gènes *MET1* sur les chromosomes des groupes 5 et 7 et de les comparer à celles du groupe 2.

	<i>MET1a</i>		<i>MET1b</i>
	TaMET1-5	TaMET1-7	TaMET1-2
Structure du gène	Altérée	Fonctionnelle	Fonctionnelle
Taux d'évolution	0,49	0,22	0,12
Cartographie	Sub-télomérique	Sub-télomérique	Peri-centromérique
Expression	-	+/-	++
CpG dans la région 5'	+++	++	-
	↓	↓	↓
	Pseudogène	?	Gène majeur

Tableau 2 : Bilan de la caractérisation des gènes *TaMET1* chez le blé tendre

Pour ce faire, il serait nécessaire de connaître les gènes adjacents par exemple sur une fenêtre de 500 à 1000kb. Ceci peut être réalisé en se basant sur les régions synténiques chez le riz et en identifiant les orthologues chez le blé grâce aux séquences génomiques IWGSC. Ces gènes seraient ensuite organisés selon l'ordre de *Brachypodium* et/ou du riz, créant ainsi une « carte virtuelle » aux alentours de *MET1*. L'étude de la vitesse d'évolution sera faite pour l'ensemble de ces gènes (calcul de dS, dN et ω), comme déjà réalisé pour les gènes *TaMET-1* dans notre publication. Les valeurs de vitesse évolutives seraient recoupées avec les données du niveau d'expression de chaque gène obtenues par l'analyse des données RNA-seq déjà utilisées. Un tel travail a été initié au laboratoire mais faute de temps n'a pas été mené à son terme.

Bilan

Le tableau 2 fait le bilan de mes travaux sur la famille *TaMET-1*. Cette étude nous indique que ce sont les gènes du groupe 2 qui devront être la cible à privilégier lors de notre recherche de mutants ou pour l'établissement de lignées RNAi. Ce sujet sera traité dans le chapitre suivant.

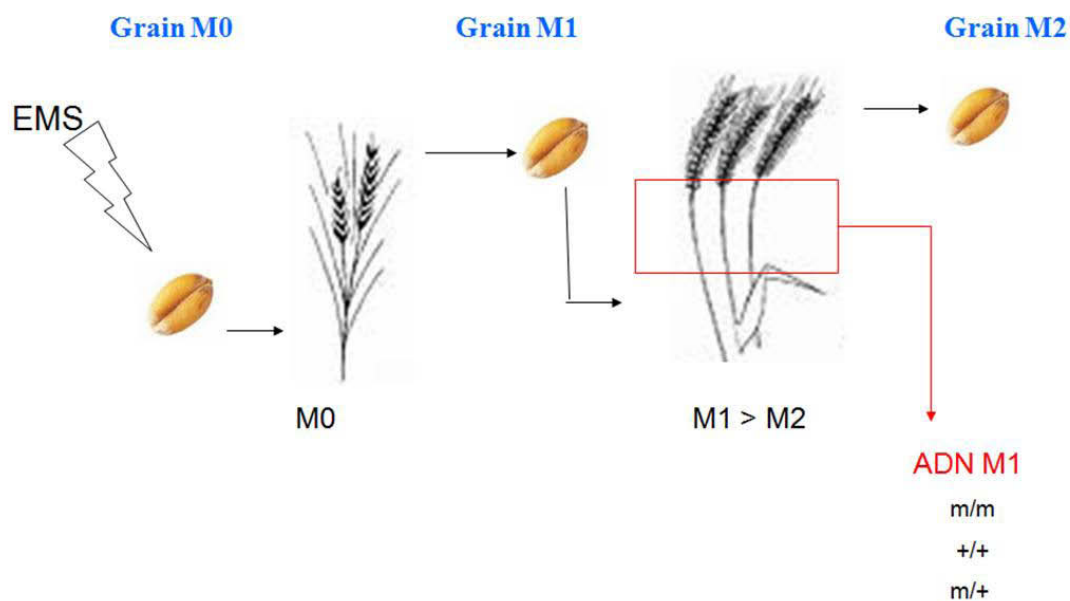


Figure 26 : Stratégie de production d'une population de TILLING

La figure détaille la nomenclature des plantes et des graines au cours des générations. Les graines M0 sont soumises à un traitement à l'EMS et semées pour obtenir des plantes M0. Des prélèvements sont effectués sur les plantes au stade M1 et regroupé avant extraction de l'ADN, il s'agit des ADN M1 testés dans les manipulations de crible. Ces plantes M1 sont également amenées à graines pour obtenir la génération M2.

Chapitre 2 : étude des profils de méthylation

Le but de cette partie de mon projet est d'étudier les conséquences d'une perte ou d'une diminution de l'activité du(es) gène(s) *TaMET-1* sur le phénotype du blé et sur les patrons de méthylation de son génome. Il s'agit donc de valider la fonction des gènes *MET1* chez le blé. La première étape a été de chercher à obtenir de telles lignées. Pour ce faire, trois stratégies classiquement utilisées en validation fonctionnelle ont été envisagées afin de mettre toutes les chances de notre côté : le crible de population de TILLING, la construction de lignées RNAi et l'utilisation de lignées aneuploïdes.

1. Lignées TILLING, RNAi, délétion

1.1. Mutants TILLING

Une première stratégie a été de rechercher des mutants de *TaMET-1* dans différentes populations de TILLING (Targeting Induced Local Lesions IN Genomes ; McCallum et al., 2000). La stratégie de production de ces populations est décrite dans la figure 26.

1.1.1. Populations de TILLING

Au départ de notre projet, la priorité était de cribler une population diploïde pour faciliter les analyses phénotypiques futures. En effet, lorsque nous avons initié la recherche de mutants, nous savions qu'au moins deux copies de *MET1* pouvaient être présentes dans un génome de blé diploïde du fait des connaissances chez d'autres céréales comme le riz ou *Brachypodium* (voir chapitre 1). Il apparaissait très ambitieux de rechercher des mutants *met1* directement chez un blé hexaploïde. En effet, cela supposait d'une part d'identifier des mutants pour chacun des génomes homéologues (A, B et D), puis dans le cas de redondance de fonction de combiner les mutations au sein d'une même plante. Le fait d'étudier un mutant *met1* dans une espèce diploïde permettait de s'affranchir des copies des autres génomes, et d'obtenir peut être des résultats plus rapidement. La recherche de mutants dans des populations hexaploïdes a tout de même été initiée.

Variété	Arche	Apache
Dose EMS	De 0,6 à 1,2%	
Graines M1	1373	5449
ADN M1	912	5350

Tableau 3 : Effectif des populations de TILLING chez le blé tendre hexaploïde disponibles à Biogemma

Nous avons à disposition deux populations hexaploïdes et une population diploïde établies par traitement à l'EMS (éthyl methane sulfonate). L'EMS induit des mutations ponctuelles en réagissant avec les guanines, menant à des transitions de type cytosine en thymine ou guanine en adénine. La première population criblée a été une population diploïde appartenant à l'INRA et établit sur *Triticum monococcum* variété *Pays de Sault*. Cette population compte 1 840 individus. La qualité de cette population semble satisfaisante puisque le pourcentage de mortalité à l'issue du traitement à l'EMS a été estimé à 20% et la stérilité à 10% (Véronique Says-Lesage, INRA, communication personnelle). Les deux populations hexaploïdes appartiennent à Biogemma. Elles ont été établies sur un blé hexaploïde des variétés *Arche* et *Apache*, qui sont respectivement un blé de printemps et d'hiver. Ces populations ont été obtenues par traitement de grains avec différentes doses d'EMS allant de 0.6 à 1.2%. La population *Arche* compte 912 ADN M1 et la population *Apache* 5 350. Les pourcentages de stérilité et de létalité ont été définis pour les deux populations en fonction de la dose d'EMS utilisée (ils varient de 20-75% en fonction de la dose d'EMS) et laissent supposer que ces deux populations contiennent un nombre important de mutations (Tableau 3 ; données confidentielles).

1.1.2.Stratégie de crible des populations

Les mutations qui apparaissent après un traitement à l'EMS peuvent être de trois types : (i) les mutations silencieuses qui n'entraînent pas de changement d'acide aminé lors de la traduction, (ii) les mutations faux sens qui entraînent un changement d'acide aminé et enfin (iii) les mutations non-sens qui entraînent l'apparition d'un codon stop. La fréquence d'apparition de ces trois classes de mutations est de l'ordre de 30% pour les mutations silencieuse, 65% pour les faux-sens et ~5% d'apparition de codon stop (McCallum et *al.*, 2000). L'identification du premier type de mutation ne présente pas d'intérêt pour nous puisque le polymorphisme ne changera pas la fonction du gène. Les mutations faux-sens nécessitent un travail supplémentaire de caractérisation de l'acide aminé et de son effet potentiel sur la protéine et son activité. Il faut ensuite tester l'effet phénotypique sur les mutations sélectionnées, ce qui peut représenter un travail conséquent et monopoliserait un certain espace en serre. Afin de réduire le travail de phénotypage, nous nous sommes focalisé sur la recherche de mutations non-sens, provoquant l'apparition de codon stop.

Séquences	Longueur de la séquence génomique	Longueur de la séquence de transcrit	Nombre de CAA	Nombre de CAG	Nombre de TGG	Nombre de CGA
T. Monococcum	3840 nt	2659 nt	10	21	15	5
C1	4186 nt	2636 nt	10	21	14	5
C2	4650 nt	3030 nt	12	25	15	6
C3	4295 nt	2845 nt	10	22	14	6
Commun aux 4 séquences			10	21	14	5

Tableau 4 : Nombre de codons CAA, CAG, TGG et CGA pour les quatre séquences utilisées lors du premier crible des populations TILLING

Suite à l'action de l'EMS, ces quatre codons sont susceptibles de donner des codons stop : CAA->TAA; CAG/TGG ->TAG; TGG/CGA -> TGA. Seuls les codons stop commun aux quatre séquences ont été ciblés pour le crible.

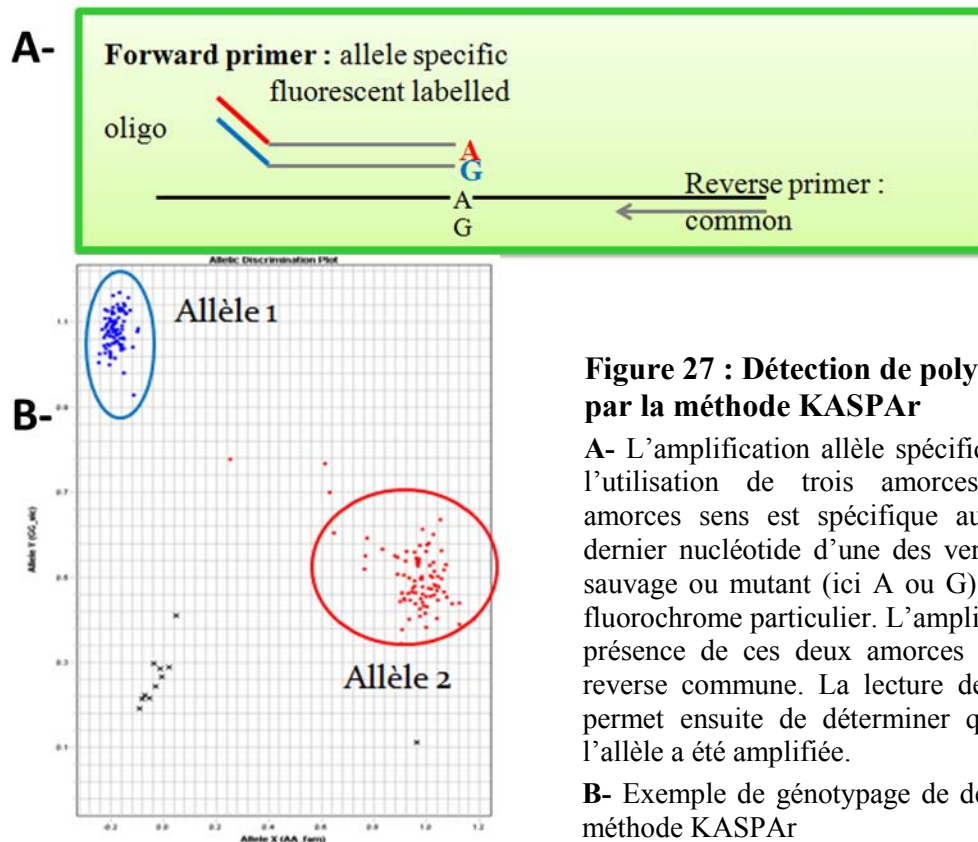


Figure 27 : Détection de polymorphisme par la méthode KASPAR

A- L'amplification allèle spécifique est basée sur l'utilisation de trois amorces. Chacune des amorces sens est spécifique au niveau de son dernier nucléotide d'une des versions de l'allèle, sauvage ou mutant (ici A ou G) et associée à un fluorochrome particulier. L'amplification se fait en présence de ces deux amorces et d'une amorce reverse commune. La lecture de la fluorescence permet ensuite de déterminer quelle version de l'allèle a été amplifiée.

B- Exemple de génotypage de deux allèles par la méthode KASPAR

Le crible des populations de TILLING a été fait en deux vagues successives. La première vague a été réalisée sur la population diploïde et sur la population hexaploïde *Arche*. Lors du lancement de cette recherche, les séquences de *TaMET-1* disponibles étaient les trois contigs (c1, c2 et c3) reconstitués lors de la première expérience de Sequence Capture ainsi que la séquence obtenue par amplification réalisée sur le diploïde *T. monoccocum* (voir chapitre 1 ; Tableau 4). La première étape a été d'obtenir la structure exons-introns afin de prédire un ARN *MET1*. Ce travail a été fait à partir d'un alignement réalisé avec l'outil GeneWise (<http://www.ebi.ac.uk/Tools/psa/genewise/>) entre les ESTs de blé (voir chapitre 1), les transcrits de *Brachypodium* correspondant à *MET1* et les contigs génomiques obtenus par Sequence Capture. La pertinence de ces jonctions exons-introns a été vérifiée par traduction des transcrits obtenus. L'ensemble des codons CAA, CAG, TGG et CGA ont été identifiés sur l'ARN *MET1* prédit et ce pour les trois contigs hexaploïdes et la séquence amplifiée chez *T. monoccocum*. La recherche de mutation s'est focalisée sur les codons stop communs aux quatre séquences, ce qui représente 50 codons stop potentiels pour l'ensemble des quatre séquences (Tableau 4). La seconde vague de crible concernait les deux populations hexaploïdes de Biogemma, elle a été lancée début 2013, et intégrée à un crible plus vaste de Biogemma auquel j'ai participé. A ce moment-là, les neuf séquences de *TaMET-1* étaient connues et les premières données d'expression disponibles. La recherche s'est donc focalisée sur la copie *TaMET-2B1* et concernait 69 codons stop.

1.1.3. Résultats du crible des populations de mutagenèse

Le crible des populations a été réalisé en utilisant la méthode KASPar développée par KBioscience. Cette méthode repose sur une amplification allèle spécifique, le polymorphisme discriminant étant placé sur le dernier nucléotide d'une des deux amorces. Dans notre cas, deux amorces allèles spécifiques sont définies grâce à l'outil PrimerPicker (KBioscience) et chacune est associée à un fluorochrome donné (Fig 27A). Il sera donc possible de savoir si l'allèle détecté est l'allèle sauvage (homozygote allèle sauvage), l'allèle mutant (homozygote allèle mutant) ou une combinaison des deux (hétérozygote) (Fig 27B).

Pour la première vague de crible sur la population diploïde et la population *Arche*, les couples d'amorces ont été définis avec PrimerPicker pour les 58 codons stop de la séquence c2 c'est-à-dire la séquence la plus longue dont nous disposions. Sur l'ensemble des amorces

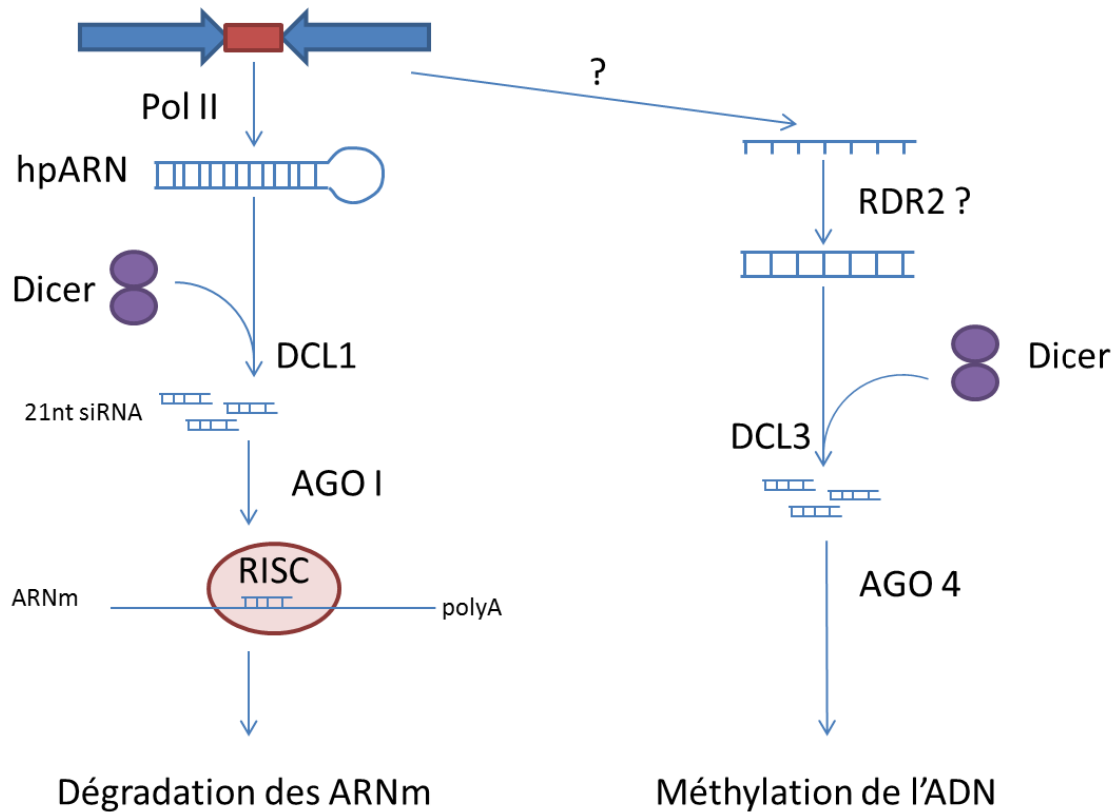


Figure 28: Principe théorique de la stratégie RNAi par formation d'ARN double brins en tige-boucle

Adapté de (Molnar *et al.*, 2011) & (Waterhouse and Helliwell, 2003)

La transcription du transgène va mener à la formation d'une structure ARN double brin (ARNdb) en tige boucle (hairpin, hpARN). Cette structure sera clivée par le complexe Dicer, composé d'un duplexe de DICER 1 (DCL1), en siRNA de 21nt. Ces petits ARNs seront pris en charge par ARGONAUTE 1 (AGO1) et le complexe RNA-induced silencing complex (RISC) menant à la dégradation des ARNm de la cible.

Dans une seconde voie possible, l'expression du transgène mènerait à la production d'un ARN aberrant, pris en charge par la RNA-DEPENDENT RNA POLYMERASE2 (RDR2) pour obtenir un ARNdb. Ce dernier serait clivé par un complexe Dicer (DCL3) puis pris en charge par AGO4 provoquant une méthylation de l'ADN cible.

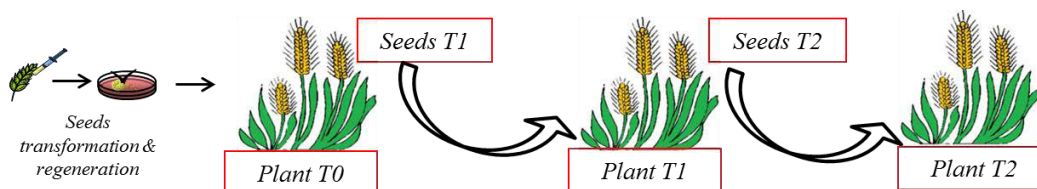


Figure 29 : Principe de transformation du blé par la Seed Inoculation Method (SIM)

Après injection de la souche d'*Agrobacterium* contenant le transgène RNAi, les embryons de blé sont régénérés par culture *in vitro*. Les cals régénérés permettront d'obtenir des plantes qui seront ensuite transférées sur terreau, il s'agit des plantes T0. Chaque cal correspond à un événement de transformation indépendant et peut être à l'origine de plusieurs plantes, appelées plantes sœurs. Les plantes T0 sont caractérisées pour le nombre de copies du transgène et l'intégrité du transgène. Les plantes sont multipliées par autofécondation et génotypées à chaque génération afin de suivre le transgène.

possible, nous avons conservé les 28 premiers couples d'amorces qui pouvaient amplifier *MET1* à la fois sur le génome diploïde et le génome hexaploïde. Les 14 premiers couples d'amorces ont été utilisés pour cette première vague de crible. Pour la deuxième vague, sur les 69 codons stop potentiels définis sur la séquence *TaMET-2B1*, les 43 premiers couples d'amorce ciblant les deux premiers exons ont été utilisés pour le crible des deux populations hexaploïdes.

Malheureusement, malgré tous nos efforts, aucun mutant n'a été identifié sur l'ensemble des populations et ce pour les deux vagues de crible. Nous ne disposons donc à ce jour d'aucun mutant EMS de *MET1*.

1.2. Lignées RNAi

Une deuxième stratégie a été de construire des lignées RNAi contre *TaMET-1*. La stratégie RNAi consiste à exprimer un transgène comportant une séquence homologue au gène à cibler. Cette séquence est organisée en répétition inversée (Fig 28) (Molnar et al., 2011 ; Waterhouse and Helliwell, 2003). Cette structure va être reconnue par la machinerie de silencing de la plante et entraîner la dégradation et/ou le blocage de la traduction de l'ARNm de *MET1*.

Cette seconde approche est sans doute plus appropriée que la recherche des mutants de TILLING. En effet, la présence probable de plusieurs copies de *MET1* suggère que si ces gènes ont gardé la même fonction, ils auront une fonction redondante et une seule copie mutée ne confèrera pas de phénotype mutant. De plus, la stratégie RNAi ne requiert pas une connaissance exhaustive de la séquence *MET1*, ce qui était notre cas au démarrage du projet. Un autre avantage est que si une réduction forte de l'expression de *TaMET-1* est létale, la stratégie RNAi permettra d'obtenir des plantes avec une inactivation partielle de *MET1* (knock down). Cette inactivation partielle permettrait de sélectionner des plantes dont le niveau d'expression de *MET1* serait suffisamment réduit pour observer une altération du méthylome sans toutefois entraîner de létalité. Suite aux premières discussions lors du lancement du projet, la stratégie RNAi est également celle privilégiée par Biogemma puisque à plus long terme l'obtention de plantes RNAi *MET1* pourrait représenter un outil important pour étudier l'épigénome du blé. Biogemma dispose d'une méthode de transformation génétique par inoculation de graine (« Seed Inoculation Method », SIM, Ishida et al., 1996) chez le blé (Fig 29). Notons toutefois qu'il n'existait pas de promoteur inductible utilisable

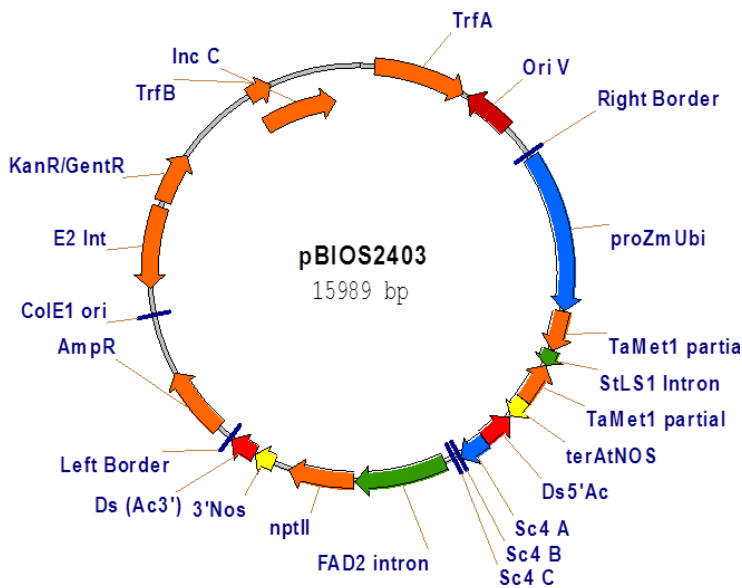


Figure 30 : Représentation schématique du vecteur utilisé pour la première vague de RNAi contre MET1

La séquence de 500pb ciblant *MET1* a été clonée afin d'obtenir deux séquences en répétition inverse. Les deux séquences sont séparées par l'intron du gène StSL1 et sous contrôle du promoteur du gène de l'ubiquitine du maïs (proZmUbi). L'arrêt de la transcription est assuré par le terminateur NOS (terAtNOS). La résistance kanamycine (KanR) permet la sélection bactérienne et le gène NptII la sélection en plante. Les séquences Right et Left Border délimitent la portion du vecteur qui sera intégrée au génome de la plante.

	Plante T0	Graines T1	Plantes T1	Graines T2
	20 plantes	14 lignées		
Transgènes intègres	3 plantes	Pour 1 lignée ↳ 14 graines ressemées	Aucune germination	
Transgènes tronqués	2 plantes	Pour les 2 lignées ↳ 15 et 16 graines ressemées	Germination pour les deux lignées T0.	Toutes les plantes germées ont été amenées à graines. Descendance T2 pour les deux lignées T0.
Transgènes négatifs	15 plantes	Pour 11 lignées ↳ 14-16 graines ressemées pour 5 de ces lignées	Germination pour les 5 lignées T0.	Toutes les plantes germées ont été amenées à graines. Descendance T2 pour les deux lignées T0.
			Prélèvements pour la recherche des siRNA	

Tableau 5 : Récapitulatif des générations de la première vague de RNAi contre MET1

20 plantes T0 ont été obtenues après transformation, mais des graines T1 n'ont été obtenues que pour 14 de ces lignées T0. Le génotypage a démontré que tous ces événements étaient mono-copie mais que le transgène était soit intègre, tronqué ou absent du criblage PCR.

La recherche de la présence de siRNA a été faite sur des prélèvements effectués sur l'ensemble des plantes au stade T1. Les résultats sont montrés en figure 29.

chez le blé qui aurait pu être utile afin de ne pas perdre d'éventuel transformant du fait d'une létalité consécutive à une forte baisse du niveau de méthylation.

1.2.1. Construction des vecteurs et clonage

Le travail sur les lignées RNAi a été démarré dès le lancement de mon projet. En effet, du fait du temps de culture du blé (10-12 mois de graine à graine), il était nécessaire de commencer au plus tôt. Au démarrage, seules les séquences des 8 ESTs identifiées et ayant servies pour définir les sondes utilisées lors du Sequence Capture étaient disponibles. La plupart des séquences correspondaient à l'extrémité 3' de *MET1*. Lors de nouvelles constructions RNAi, Biogemma cible de préférence la région 3' des gènes ce qui dans notre cas s'avère une chance puisque nous ne disposions que de cette région.

A partir de ces séquences d'ESTs, une séquence de 500pb a été définie pour établir les constructions en évitant toutefois la région 3' non traduite (3'UTR) (Fu et al., 2007) et en s'assurant de la spécificité de la région ciblée par des analyses BLASTn. Ce fragment est organisé en répétition inverse, séparé par un espaceur (intron du gène *StLS1* de la pomme de terre) et a été inséré dans un vecteur sous contrôle du promoteur du gène de l'ubiquitine de maïs connus pour s'exprimer chez le blé (données Biogemma non publiées) (Fig 30).

1.2.2. Première vague de transformants RNAi *MET1*

Après transformation des embryons de blé de la lignée NB1 par la méthode SIM, 7 évènements indépendants de génération T0 ont été récupérés. Au vu du faible nombre de transformants, des plantes sœurs c'est-à-dire des plantes issues du même cal transformé ont été récupérées pour au moins deux évènements (Tableau 10). Au total, 20 plantes T0 issues de 7 évènements indépendants ont donc été régénérées. Sur ces 20 plantes, une majorité semble avoir un transgène réarrangé (Tableau 5). Une seule des lignées positive pour le transgène a donnée des graines T1 mais aucune de ces graines n'ont germé (lignée T0 B0053041). Les graines des deux lignées avec un transgène tronqué ont été ressemées, ainsi que celles de cinq des lignées avec un transgène absent selon le génotypage PCR, dans le but d'obtenir une descendance T2.

Vu le peu de plantes obtenues en T0 et l'importante létalité et stérilité observées à chaque génération, toutes les lignées ont été conservées quelque soit l'intégrité du transgène. Une

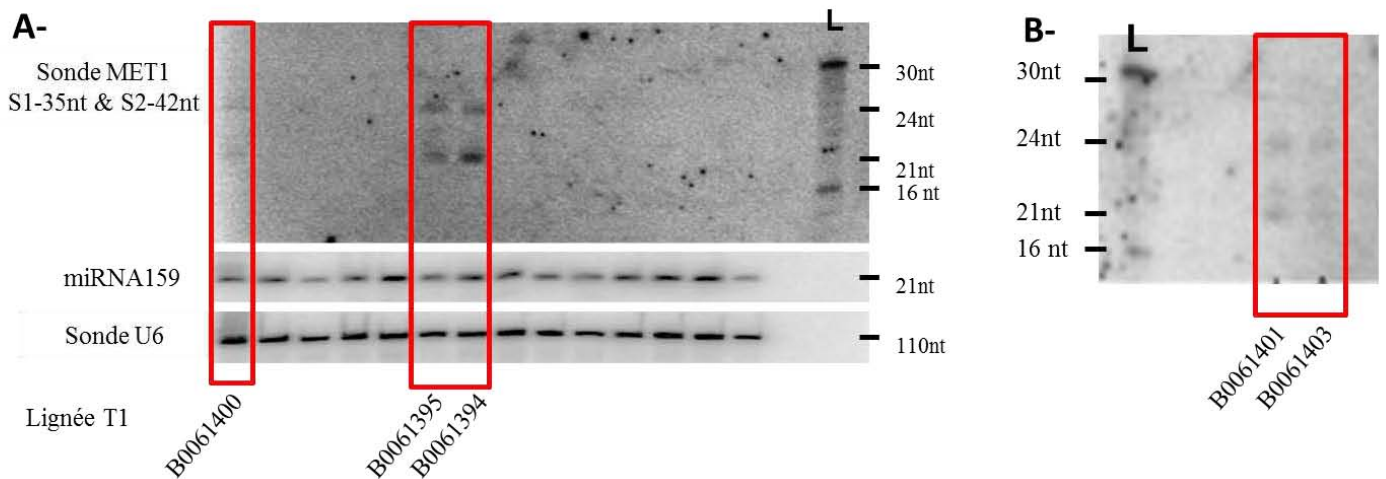


Figure 31 : Résultats de recherche de siRNA par Northern Blot dans les lignées T1 exprimant un transgène RNAi dirigé contre *MET1*

Les siRNA ciblant *MET1* ont été révélés avec un mélange des sondes 1 et 2 dirigées contre *MET1* (voir Matériels et Méthodes). Le miRNA159 a été utilisé comme marqueur de taille de 21nt.

L : échelle externe d'oligonucléotides de 16, 21, 25 et 30nt reconnus par les sondes 1 et 2 dirigées contre *MET1*.

A- sur le premier gel, trois échantillons montrent la présence de siRNA (encadré en rouge).

B- sur le second gel deux échantillons montrent la présence de siRNA (encadré en rouge). Les hybridations avec les sondes contre le miRNA159 et l'U6 ne sont pas disponibles pour ce second Northern Blot.

Lignée T0	Plantes T1	Génotypage	Production de siRNA	Nombre de graines T2	1 ^{er} semi	Germ ^o semi 1	2 ^{ème} semi	Germ ^o semi 2
B0053038	B0061393	Hz		12				
	B0061394	Hz	✓	13	5	1	8	6
	B0061395	Hz	✓	12	5	1	7	6
	B0061396	NS		29				
	B0061397	NS		42				
	B0061398	NS		40				
	B0061399	NS		14				
	B0061400	NS	✓	//				
	B0061401	He	✓	28	10	8	5	3
	B0061402	He		6				
	B0061403	He	✓	26	10	8	5	3
	B0061404	He		8				

Tableau 6 : Récapitulatif des semis des plantes T2 issus de la lignée T0 B0053038

Ce tableau reprend l'ensemble des informations sur les plantes T1 issus de la lignée T0 B0053038 : leur génotypage, la production de siRNA et les semis réalisés pour les analyses à la génération T2.

descendance T2 est disponible pour l'ensemble des lignées ayant survécues, le nombre de grains disponible est toutefois très variable selon les lignées (Matériels et Méthodes).

1.2.3. Recherche de siRNA dirigés contre *MET1*

Au stade T1, à 1 mois et demi et à 2 mois après le semi, des échantillons ont été prélevés pour vérifier la production de petits ARNs (siRNA) par Northern Blot en utilisant un mélange des sondes oligonucléotidiques (S1-35nt et S2-42nt) ciblant les siRNA *MET1*. Seules 5 plantes T1 montrent la production de siARNs dont la taille est approximativement de 21 et 24 nt (Fig 31). De façon satisfaisante, les 5 plantes T1 dérivent toutes de la même lignée T0 (B0053038, Tableau 6). Cet évènement comporterait selon les analyses PCR réalisées par l'équipe de transgénèse de Biogemma un construit tronqué au niveau du promoteur ubiquitine. Quatre de ces plantes T1 positives ont donné des graines T2.

Les résultats de cette première vague de transgénèse sont décevants. La majorité de nos évènements ont été perdus au cours des croisements successifs et seul un évènement semble produire des siRNA. L'observation de deux types de siRNA pourrait s'expliquer par des siRNA de 21 nucléotides (nt) compétents pour déclencher le PTGS auxquels s'ajouteraient des *trans-acting* siRNA de 24nt impliqués dans la diffusion systémique du signal de silencing (Vaucheret, 2005). Une autre hypothèse pourrait être que les siRNA de 24nt seraient produit du fait de l'hétérochromatinisation du transgène. En effet, si notre transgène est réarrangé il pourrait être reconnu comme une séquence répétée et serait la cible du RdDM dont l'un des médiateurs est les siRNA de 24nt. Il est assez courant dans ce cas de figure de retrouver également des siRNA de 21nt témoins de la production d'ARN aberrants dégradés par PTGS. Quelque soit le mécanisme mis en jeu, nous avons considéré que les siRNA produits pourraient également cibler les ARN de *TaMET-1* (mécanisme dit de co-suppression) et avons donc poursuivi nos travaux avec la seule lignée produisant des siRNA.

1.2.4. Analyses phénotypiques des lignées RNAi

Pour la suite des expérimentations, nous avons utilisé les graines T2 issues des plantes T1 produisant des siRNA (évènement T0 B0053038 – Tableau 6). Un premier semi a été fait début juillet 2013 à la serre de Biogemma en effectuant un semi de 10 graines issues des hétérozygotes (B0061401 & B0061403) et 5 graines issues des homozygotes (B0061394 & B0061395). A partir des données de génotypage réalisées sur les plantes T2, certaines ont été

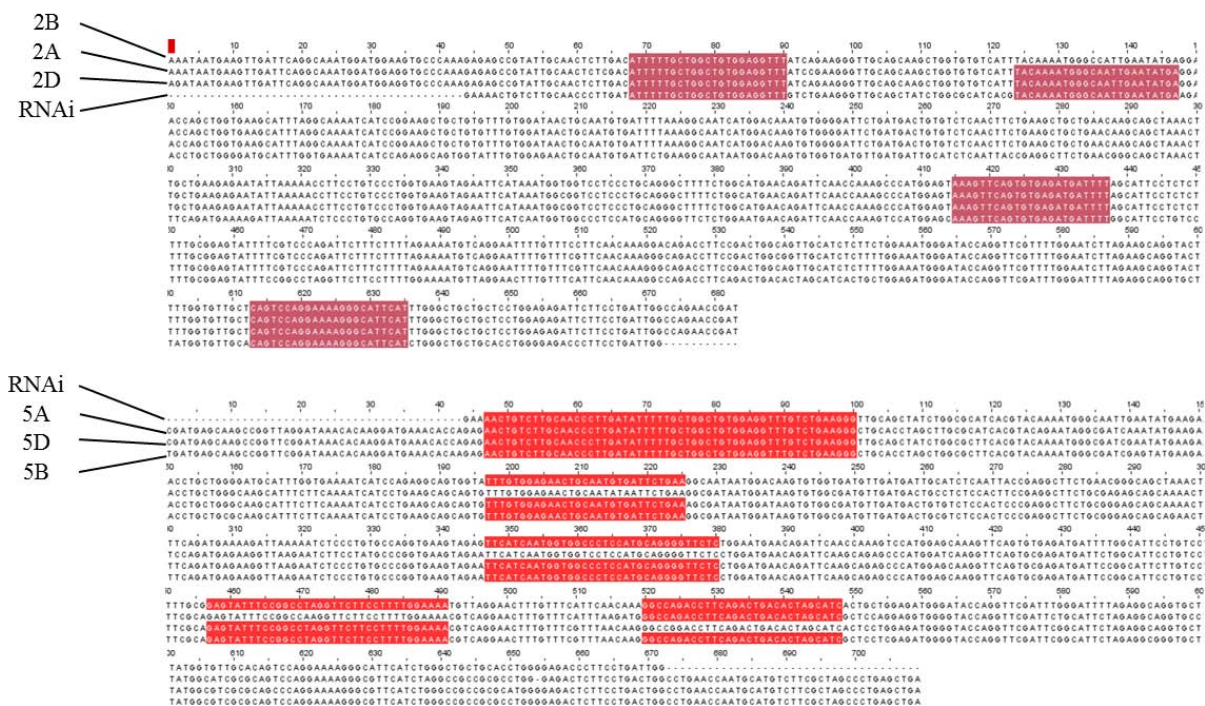


Figure 32 : Alignement de la séquence incluse dans le transgène RNAi « vague 2 » et les transcrits *TaMET1* des groupes 2 et 5.

La séquence RNAi dérive de la séquence *TaMET1* du groupe 7. Les zones en couleurs représentent les régions d'au moins 21 nucléotides identiques entre la séquence RNAi et les séquences des groupes 2 et 5.

	Plante T0	Graines T1	Plantes T1	Graines T2
	12 plantes	12 lignées		
Transgènes intègres	8 plantes	Pour les 8 lignées ↪ entre 6 et 16 graines ressemées	Germination pour toutes les lignées T0	Pour 5 des lignées T0, 2He et 2Hz et pour les 3 restantes, 1Hz et 2He amenées à graines.
Transgènes tronqués	1 plantes	Pour la lignée ↪ 16 graines ressemées	Germination pour la lignées T0.	2Hz et 2He amené à graines
Transgènes négatifs	3 plantes	Pour les 3 lignées ↪ entre 14-16 graines ressemées	Germination pour toutes les lignées T0. Prélèvements pour analyse de la méthylation de l'ADN par HPLC	Plantes éliminées

Tableau 7 : Récapitulatif des générations de la deuxième vague de RNAi contre *TaMET1*

12 plantes T0 indépendantes ont été obtenues. Le transgène est soit intègre, tronqué ou absent du crible PCR. Les lignées ne présentant pas de transgène ont été éliminées au stade plante T1.

Pour chacune des lignées T0 restantes, 1 à 2 plantes T1 homozygotes (Hz) et 2 plantes T1 hétérozygotes (He) sont amenées au stade graines T2.

Des prélèvements sur les plantes T1 ont été faits afin de réaliser une analyse globale de la méthylation par HPLC.

cultivées afin de produire des graines : 1 nul ségrégant (absence du transgène ; NS), 2 hétérozygote (He) et 1 homozygote (Hz) pour les deux lignées T1 B0061401 et B0061403 (soit 8 plantes) (Tableau 6). Les autres plantes ont été prélevées au stade Z41-45 sur l'échelle de Zadok (Zadoks et *al.*, 1974) soit environ 2,5 mois après le semis. Ces premiers échantillons ont été perdus. En effet, ils ont malheureusement été retrouvés à l'extérieur du congélateur - 80°C fin septembre.

Comme les 8 plantes conservées dans l'expérimentation ci-dessus (pour la production de graines) étaient trop avancées pour effectuer de nouveaux prélèvements de bonne qualité, un second semis a donc été réalisé début octobre à l'INRA de Crouël. Il n'était en effet pas possible d'obtenir de l'espace en serre à Biogemma. Au final, nous avons réussi à semer une seconde fois 5 graines issues des deux lignées T1 hétérozygotes (B0061401 & B0061403) et le restant des graines pour la descendance des plantes T1 homozygotes (B0061394 & B0061395) (Tableau 6).

Suite à ces deux semis, il ne reste aucune graine T2 issues des lignées T1 homozygotes productrices de siRNA.

1.2.5. Seconde vague de transformants RNAi *MET1*

Au cours de l'année 2012, Biogemma s'est rendu compte qu'il y avait une erreur dans le construit RNAi au niveau de la séquence comprise entre le promoteur ubiquitine et l'intron. Ceci entraînerait une mauvaise efficacité de l'épissage et donc limiterait l'expression de la construction RNAi et/ou la stabilité des ARNs produits. Toutefois, il convient de remarquer qu'au moins pour un évènement, les résultats de Northern Blot suggèrent que des siRNA peuvent être produits. Le problème d'épissage pourrait être peu impactant pour nos construits.

Une deuxième vague de RNAi a donc été lancée malgré le fait que les lignées qui seraient établies ne seraient sans doute pas analysables au cours de mon projet de thèse. A cette époque (fin 2012-début 2013) les 9 séquences de *TaMET-1* étaient connues. Ceci a permis de choisir une séquence qui présentait des identités de séquence avec les 9 copies et donc d'obtenir un construit RNAi susceptible de cibler toutes les copies de *MET1*.

La séquence utilisée a été construite de la même manière que la précédente en privilégiant la fin du gène et en vérifiant la spécificité contre *MET1* en réalisant des analyses BLASTn contre l'ensemble des séquences génomiques IWGSC. La séquence ciblant *MET1*

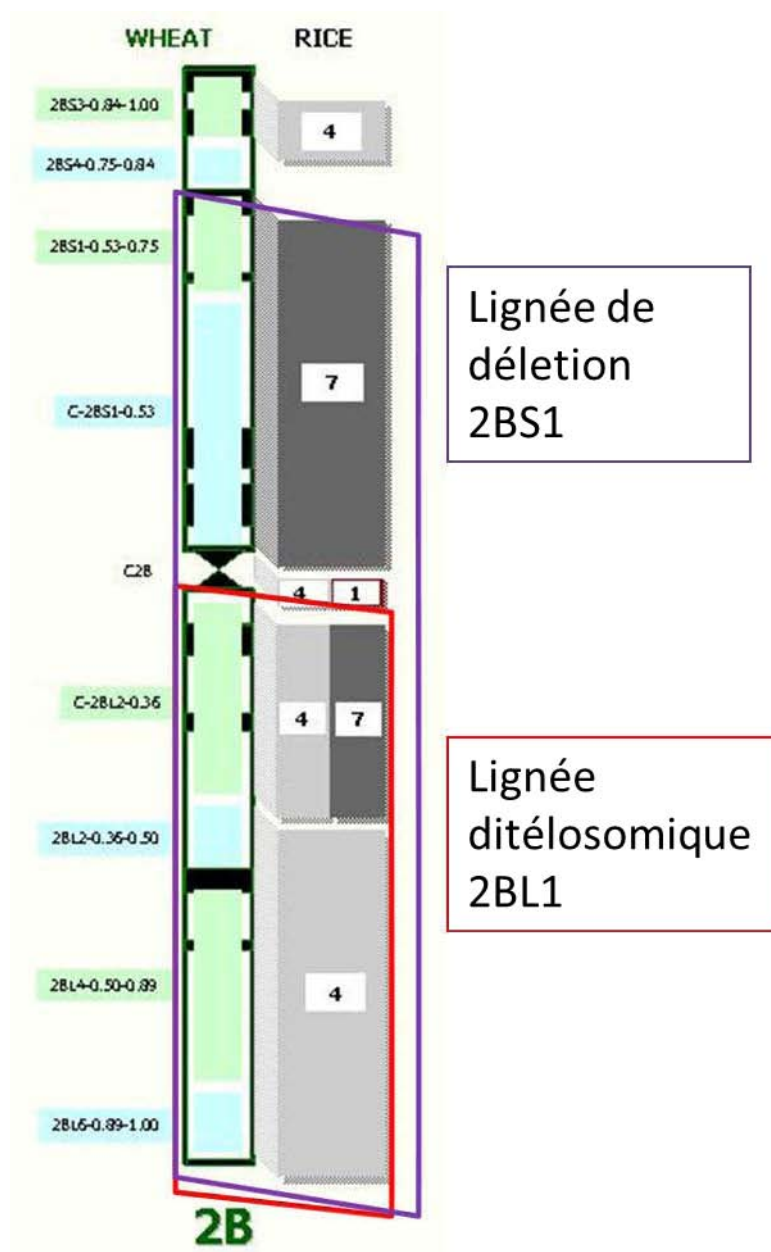


Figure 33 : Représentation schématique des lignées aneuploïdes utilisées

Adapté de http://wheat.pw.usda.gov/wEST/binmaps/wheat2_rice.html

Les régions vertes et bleues représentent les différents blocs de délétions manquants dans les différentes lignées affectant le chromosome 2B. La synténie avec le riz est indiquée en gris pour les différentes régions.

La partie encadrée en violet correspond au fragment de chromosome présent dans la lignée de délétion 2BS1, et celle encadrée en rouge à la portion présente dans la lignée ditélosomique 2BL1.

utilisée dans cette nouvelle construction compte 630pb et d'après l'analyse BLASTn cette séquence permettrait de cibler les 9 copies avec la même construction RNAi (Fig 32).

22 évènements T0 indépendants ont été obtenus mais comme lors de la première vague, une forte stérilité a été observée et seules 12 plantes ont donné des descendants (Tableau 7). Sur ces 12 plantes T0, 8 sont positives pour l'insertion d'un transgène intègre, une plante possède un transgène tronqué et les trois plantes restantes sont négatives pour la présence du transgène. L'ensemble de ces plantes a tout de même été mené au stade graines T1, une partie ayant été ressemée (voir Matériels et Méthodes). Après génotypage de l'ensemble des plantes T1 germées, celles issus des lignées T0 sans transgène ont été éliminées et 1 à 2 homozygotes et 2 hétérozygotes ont été conservés pour les autres lignées pour produire des graines.

Cette seconde vague est plus prometteuse puisque 8 évènements sont monocopies pour le transgène qui ne semble pas avoir subi de réarrangement. Des échantillons seront envoyés fin Avril début Mai 2014 à Stéphane Maury (Université d'Orléans) pour analyse du taux global de méthylation par une méthode basée sur une analyse HPLC des nucléosides (Trap-Gentil et *al.*, 2011). Nous envisageons également de vérifier la production des siRNA dans les plantes issues des 8 évènements d'intégration du transgène par Northern Blot.

1.3. Lignées de délétion

Notre troisième stratégie pour obtenir des lignées ayant un profil de méthylation modifié est basée sur l'étude de lignées aneuploïdes. Ces lignées initialement isolées par Endo et Gill (1996) sont disponibles auprès du centre de ressources génétiques de l'INRA de Crouël où elles ont été amplifiées. Il est possible de travailler avec des lignées nullisomique-tétrasoniques, c'est-à-dire pour lesquelles il manque une paire de chromosomes remplacée par une paire de chromosomes homéologues ; des lignées ditélosomiques, pour lesquelles il manque les deux copies homéologues d'un des bras chromosomique ; et enfin des lignées de délétion pour lesquelles il manque une partie plus ou moins importante d'un des bras chromosomique (Saintenac, 2009). Ces lignées nous ont été fournies par Pierre Sourdille (INRA Crouël).

Les résultats d'analyse de l'expression des 9 copies de *TaMET-1* par RNAseq (voir chapitre 1) ont montré que ce sont les copies présentes sur les chromosomes 2 qui sont les plus exprimées. Les analyses complémentaires réalisées par RT-qPCR ont confirmé ces résultats et suggèrent une expression un peu plus importante de la copie *TaMET-2B1*. Ces

		Plante T0	Graines T1	Plantes T1	Graines T2	Plantes T2
Lignées RNAi	RNAi vague 1	20 plantes	14 lignées dont 1 avec transgène intègre	7 lignées dont 1 avec des siRNA	1 lignée sélectionnée	Prélèvements à analyser
	RNAi vague 2	12 plantes	Pour toutes les lignées, dont 8 avec le transgène intègre	12 lignées, mais élimination des 3 lignées négatives pour le transgène	En cours pour les 9 lignées avec transgènes positives ou tronqué	
Populations de TILLING	TILLING diploïde (T. monococcum)	1200 ADN criblés	Aucun mutant pour 14 codons stop recherchés			
	TILLING hexaploïde (Arche)	912 ADN criblés	Aucun mutant pour 14 + 43 codons stop recherchés			
	TILLING hexaploïde (Apache)	5350 ADN criblés	Aucun mutant pour 43 codons stop recherchés			
Lignées de délétion (CHS)		2BS1, DT2BL	En cours			
Mutants d'irradiation (Renan)		4000 ADN criblés	Candidats en cours de confirmation			

Tableau 8 : Récapitulatif du matériel végétal criblé (TILLING, lignées de délétion et mutants d'irradiation) ou produit (RNAi) au cours de mon projet de thèse.

copies du groupe 2 étant cartographiées sur le bras court (2BS), nous avons utilisé une lignée de délétion du 2BS appelée « 2BS1 », qui contient encore le segment chromosomique qui porterait d'après nos données de cartographie le gène *TaMET-2B1*; et la lignée ditélosomique 2BL1 qui elle serait délétée pour le locus *TaMET-2B1* (Fig 33).

Les semis de ces lignées ont été faits en même temps que les semis des lignées RNAi. Le premier semis de 5 graines de chaque lignée et 5 graines du témoin *T. aestivum* var. *Chinese Spring* a été perdu. Le deuxième semis a été effectué avec les mêmes effectifs à l'INRA de Crouël. Ces échantillons sont en cours d'analyses.

1.4. Bilan de l'obtention de lignées dérégulées pour *TaMET-1*

Au final, nous disposons de deux vagues de lignées RNAi, la première avec un doute sur la fonctionnalité du transgène qui nous a permis d'obtenir 7 évènements indépendants dont 1 lignée produit des siRNA contre *TaMET-1*. La seconde vague a permis d'obtenir 8 évènements indépendants en cours d'analyse. Aucun mutant EMS n'a pu être identifié. Nous disposons également d'une lignée ditélosomique dépourvue de la copie *TaMET-2B1*, c'est-à-dire délétée de la copie qui semble la plus exprimée. Notons que des recherches sont en cours afin d'identifier des lignées de délétions dans la population d'irradiation *T. aestivum* var. *Renan*. Ce travail est actuellement réalisé par l'équipe de Pierre Sourdille à l'INRA de Crouël. L'ensemble de ces stratégies et leurs résultats est résumé dans le tableau 8.

L'ensemble de ces lignées pourraient présenter potentiellement une modification de leur profil de méthylation. Pour vérifier si tel est le cas, nous avons choisi une étude des profils de méthylation par la méthode de conversion au bisulfite (Herman et al., 1996) en étudiant des « gènes candidats » tel que *Vrn-1* connus pour présenter des régions fortement méthylées en contexte CpG (Khan et al., 2013). L'approche d'étude de la conversion au bisulfite sur des gènes candidats sera décrite dans la partie suivante.

2. Etude des profils de méthylation

2.1. Profils de méthylation de gènes candidats

La mise au point de la conversion au bisulfite a été décrite dans le chapitre précédent. Cette méthode a été utilisée afin d'analyser les profils de méthylation dans les plantes

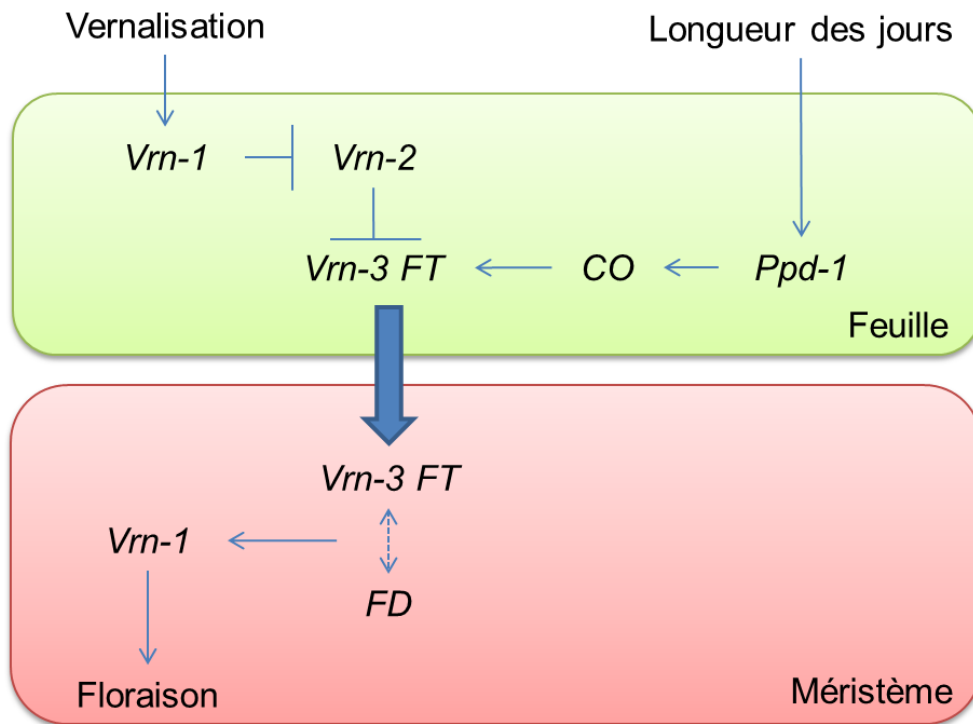


Figure 34 : Modèle d'interaction des gènes impliqués dans la floraison du blé

Adapté de Ream et *al.*,2012 et Trevaskis et *al.*,2007

La transition florale est un caractère agronomique très utilisé par le sélectionneur pour adapter des variétés à des localisations géographiques différentes (Nord / Sud). La floraison est une composante importante du rendement en grain.



Figure 35 : Effet de plusieurs allèles du gène *Rht-1* (nanisme) sur la taille des tiges chez le blé

Source : Pearce et al 2011.

Le nanisme est un caractère agronomique important car plus une variété sera haute plus elle aura tendance à verser. La verse induit des pertes importantes de rendement en grain.

exprimant un RNAi contre *MET1* ou présentant une délétion de la région chromosomique contenant *TaMET-2B1* (voir paragraphe 1.2. & 1.3.). Nous souhaitons étudier les profils de méthylation de quelques gènes préalablement choisis avec Biogemma ainsi que les régions en amont des gènes *MET1*.

2.1.1. Choix des gènes candidats

2.1.1.1. Gènes du rendement

Les premiers gènes candidats choisis ont été les trois amplicons *Vrn-1* sur lesquels la mise au point de la méthode de conversion au bisulfite a été faite. En effet, le profil de ces trois amplicons est connu : deux présentent un profil non méthylé (*VRN1-1* et *VRN1-2*) alors que la troisième région (*VRN1-3*) est fortement méthylée en contexte CpG (Khan et al 2013). Cette dernière région est particulièrement appropriée pour notre étude puisqu'elle est la cible du principal contexte de méthylation maintenu par *MET1* (contexte CpG). Ces amplicons ayant été utilisés pour la mise au point de la conversion au bisulfite, ils seront donc de bons témoins de l'efficacité de conversion des cytosines en uracile. *Vrn-1* est un gène impliqué dans la vernalisation. La vernalisation est définie comme le processus qui permet à la plante de passer de l'état végétatif à l'état reproductif après une exposition au froid (Fig 34) (Ream et al., 2012 ; Trevaskis et al., 2007). Ce processus de transition florale est important car il permet aux plantes de fleurir dans des conditions déterminées de température et de durée de jour. La question du temps de floraison est d'intérêt pour Biogemma car elle représente une des composantes impliquées dans l'établissement du rendement. En plus de la vernalisation, nous avons également retenu la longueur du jour qui est un autre facteur déterminant la transition florale (Fig 31) (Pearce et al., 2011). Trois gènes de cette voie ont été sélectionnés : *VERNALIZATION 2* (*Vrn-2*), *VERNALIZATION 3* (*Vrn-3*), également connu comme *FLOWERING LOCUS T* (*FT*), et le gène *PHOTOPERIOD 1* (*Ppd-1*).

Un dernier gène a été ajouté à l'étude, le gène *REDUCED HEIGHT* (*Rht1*). Il a été montré que l'introduction de ce gène chez le blé a permis une augmentation importante des rendements. En effet, ceci a permis de diminuer la taille des tiges (Fig 35) et donc de limiter l'effet de verse (Hedden, 2003 ; Pearce et al., 2011).

Au final, notre choix s'est porté sur les gènes suivant : *Ppd-1*, *Vrn-1*, *Vrn-2*, *Vrn-3*, *Rht-1* et la région promotrice de *Rht-A1*.

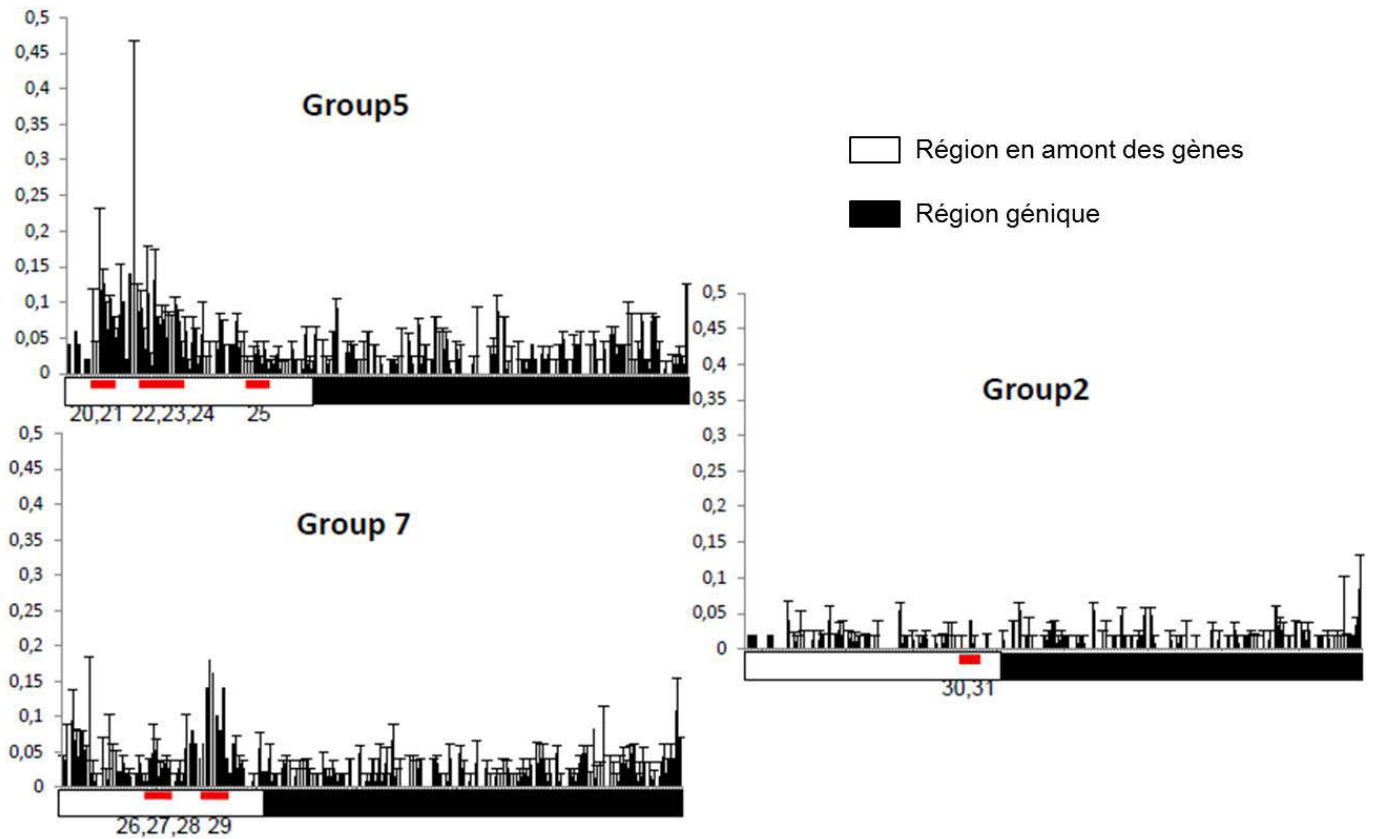


Figure 36 : Récapitulatif des régions promotrices *TaMET1* choisies pour l'étude des profils de méthylation

Les profils de CpG ont été établis à partir des alignements des trois séquences homologues pour chacun des groupes (voir chapitre 1). L'axe des ordonnées correspond au nombre de CpG pour 50 bases. En abscisse, le rectangle blanc correspond à la partie en amont du cadre de lecture et incluant le promoteur et la partie noire la partie codante. Chaque région étudiée est nommée par un numéro et est indiquée par un trait rouge. Après traitement de l'ADN génomique au bisulfite, chaque région sera amplifiée par PCR puis séquencée.

2.1.1.2. Gènes *TaMET-1*

Parmi les gènes candidats, nous avons également choisi d'inclure les gènes *TaMET-1*, en particulier les régions en amont de la région codante déjà étudiées dans le chapitre 1. Des amorces sont déjà disponibles au niveau de ces régions promotrices et les profils de méthylation des régions amplifiées sont connus dans un contexte sauvage (voir article 1). De plus, la région en amont de la copie *TaMET-7DI* présente une région très riche en CpG, cette région sera donc très intéressante à étudier dans les contextes mutants au même titre que la région couverte par l'amplicon *VRN1-3*.

2.1.2. Choix des régions géniques étudiées

Une fois la liste des gènes établie, il nous a fallu définir la ou les régions qui seraient étudiées. Ces régions doivent au minimum couvrir des zones comportant des sites méthylables ou potentiellement méthylables dans le contexte CpG.

Pour les régions d'étude sur les gènes *TaMET-1*, nous avons réutilisé les amplicons spécifiques des groupes 5 et 7 définis dans le chapitre 1 (voir article 1). Trois couples disponibles mais non exploités dans le cadre de notre publication ont été ajoutés pour le groupe 5 et de nouveaux couples ont été définis dans la région en amont des copies du groupe 2, non étudié dans notre publication. La figure 36 reprend tous les couples qui seront étudiés dans le cadre de ce chapitre.

A partir des séquences des gènes *Ppd-1*, *Vrn-1*, *Vrn-2*, *Vrn-3*, *Rht-1* et la région promotrice de *Rht-A1* fournies par Biogemma et complétées pour certaines par des séquences IWGSC, une première recherche de régions d'intérêt a été faite avec des enzymes de restriction sensibles à la méthylation. Le but était de déterminer des zones à privilégier lors de l'étude de la méthylation par conversion au bisulfite. Mon choix s'est porté sur McrBC qui est une enzyme qui coupe lorsque la séquence est méthylée et sur le couple d'isoschisme HpaII/MspI sensibles à la méthylation (Fig 37). Cette recherche a été faite pour toutes les séquences de gènes cibles sauf pour *Vrn-1*. Suite à des problèmes de mise au point des amplifications PCR, aucun résultat n'a pu être obtenu pour les séquences des gènes *Rht-A1* et pour la région promotrice *Rht-A1*. Un exemple des profils d'amplification obtenus après digestion est montré pour le gène *Ppd-1* dans la figure 37 et l'ensemble des résultats pour ce même gène est montré dans la figure 38.

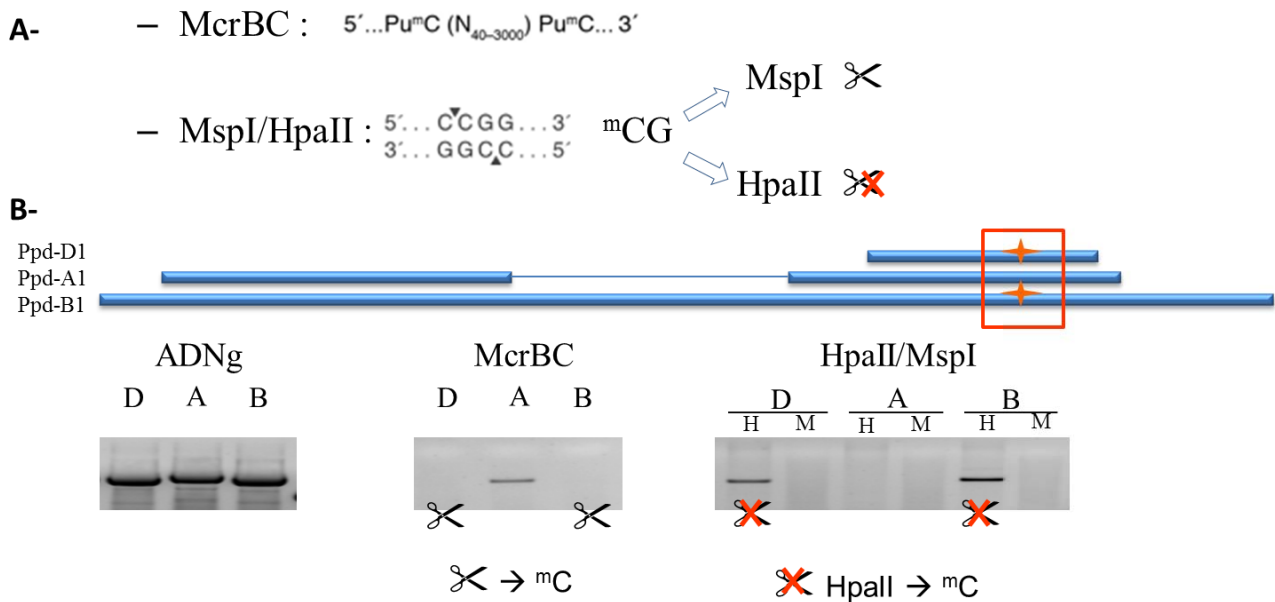


Figure 37 : Principe du crible d'identification de régions méthylées par digestion avec des enzymes sensibles à la méthylation

A- L'enzyme McrBC est capable de couper l'ADN entre deux cytosines méthylées lorsqu'elles sont séparées par au moins 40pb selon le motif indiqué. Les enzymes HpaII (H) et MspI (M) reconnaissent la même séquence, mais la coupure avec HpaII est bloquée si la méthylation intervient en contexte CpG.

B- Exemple de résultats obtenus après utilisation de ces enzymes sur le gène *Ppd-1*. Après coupure avec McrBC ou avec le couple MspI/HpaII, trois amplicons homéologues spécifiques couvrant la même région du gène *Ppd-1* ont été amplifiés. L'absence de bande pour les copies D et B après coupure avec McrBC montre que ces deux copies sont méthylées. L'absence de bande dans la piste M alors qu'elle est présente dans la piste H pour ces mêmes copies, confirme la présence de méthylation. Les résultats sont résumés dans l'alignement schématique de ces trois copies, les méthylations signalées par les étoiles rouges.

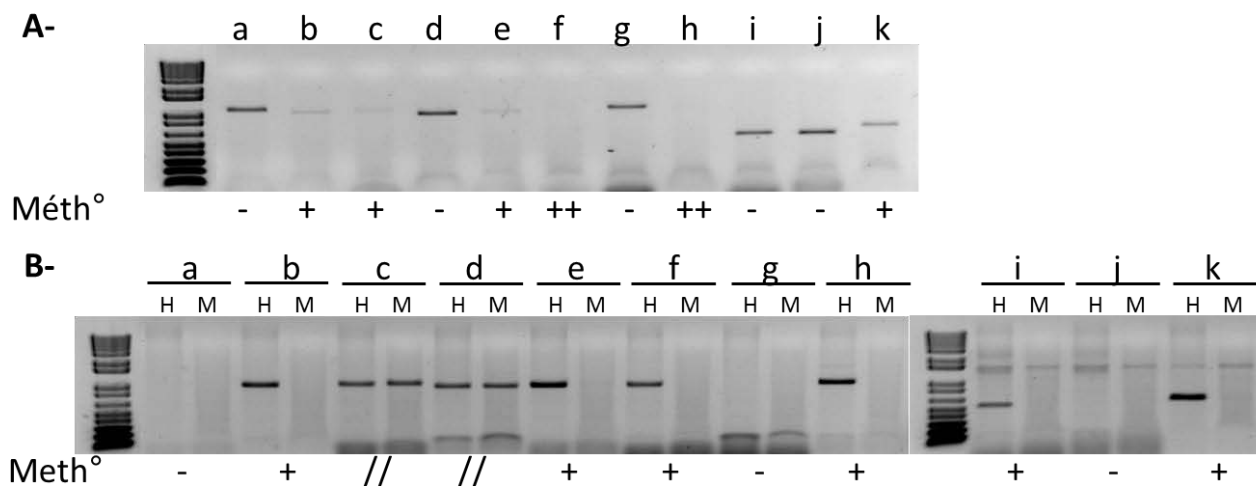


Figure 38 : Résultats du crible d'identification de régions méthylées par digestion par des enzymes sensibles à la méthylation pour le gène *Ppd-1*

Les lettres a à k désignent les différents amplicons étudiés sur le gène *Ppd-1*

A- amplification après coupure avec McrBC

B- amplification après coupure avec HpaII (H) et MspI (M)

Les amplicons b, e, f, h et k montrent la présence de méthylation avec les deux méthodes

. - : absence de méthylation ; + : présence de méthylation ; // : pas de site de reconnaissance des enzymes

Finalement, le choix des zones à étudier a été fait en prenant en compte trois critères :

- la région amplifiée doit-être présente sur les trois génomes homéologues, bien que ceci n'ait pas été possible pour tous les gènes.
- la région amplifiée doit contenir des CpG selon la prédiction réalisée par le script perl utilisé dans le chapitre 1.
- la région doit être méthylée selon notre crible préliminaire effectué avec les enzymes de restriction sensibles à la méthylation.

La figure 39 (A et B) résume les résultats des cribles enzymatiques, la présence de site CpG potentiellement méthylables et la localisation des amplicons sur le gène d'intérêt. Nous disposons de 34 régions analysables répartis dans 7 groupes de gènes candidats: 5 couples pour *Vrn-1*, 12 couples pour *TaMET-1*, 4 couples d'amorces pour *Ppd-1*, 2 couples pour *Vrn-2*, 3 couples pour *Rht*, 3 couples pour la région promotrice de *Rht-1* et 5 couples pour *Vrn-3* (voir Matériels et Méthodes).

2.2. Séquençage MiSeq

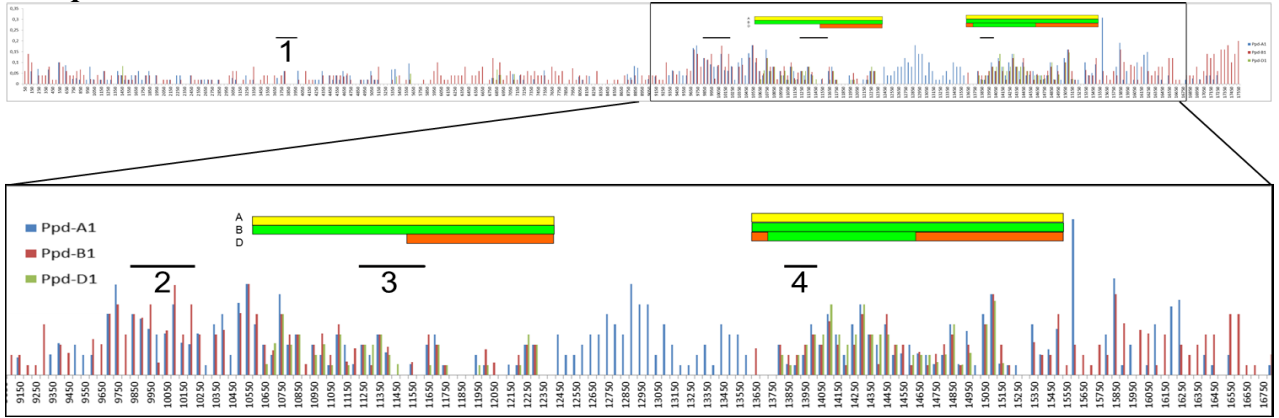
Les 34 amplicons définis ci-dessus seront amplifiés sur les 8 génotypes choisis : 5 lignées RNAi, dont la lignée témoin NB1 utilisée pour la transgénèse à Biogemma, 2 lignées aneuploïdes et le génotype sauvage dont elles sont issues (*T. aestivum* var. *Chinese Spring*). Après extraction d'ADN, traitement au bisulfite et amplification PCR, les fragments obtenus seront séquencés par séquençage haut débit de type MiSeq (Illumina).

2.2.1. Mise au point

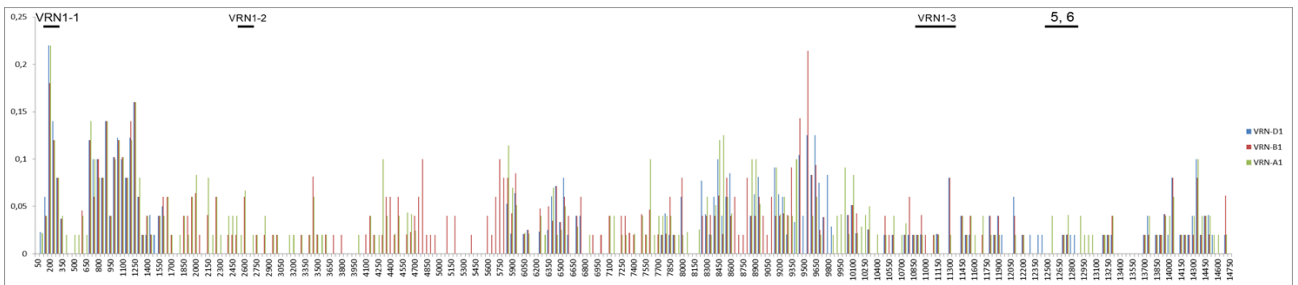
2.2.1.1. Technique de séquençage : MiSeq

Plusieurs techniques de séquençage ont été envisagées. La technique de référence de Sanger (Sanger et al., 1977) a rapidement été écartée du fait du cout important (34 couples x 8 génotypes x 20 clones x 5€ / séquence, soit près de 27k€) et de la nécessité de réaliser au préalable le clonage des segments d'ADN à séquencer. Cette technique pourra cependant être utilisée dans des phases préliminaires de validation. Nous avons initialement travaillé avec la technique 454 de Roche (Margulies et al., 2005) permettant un séquençage d'amplicons de

Ppd-1



Vrn-1



Vrn-2

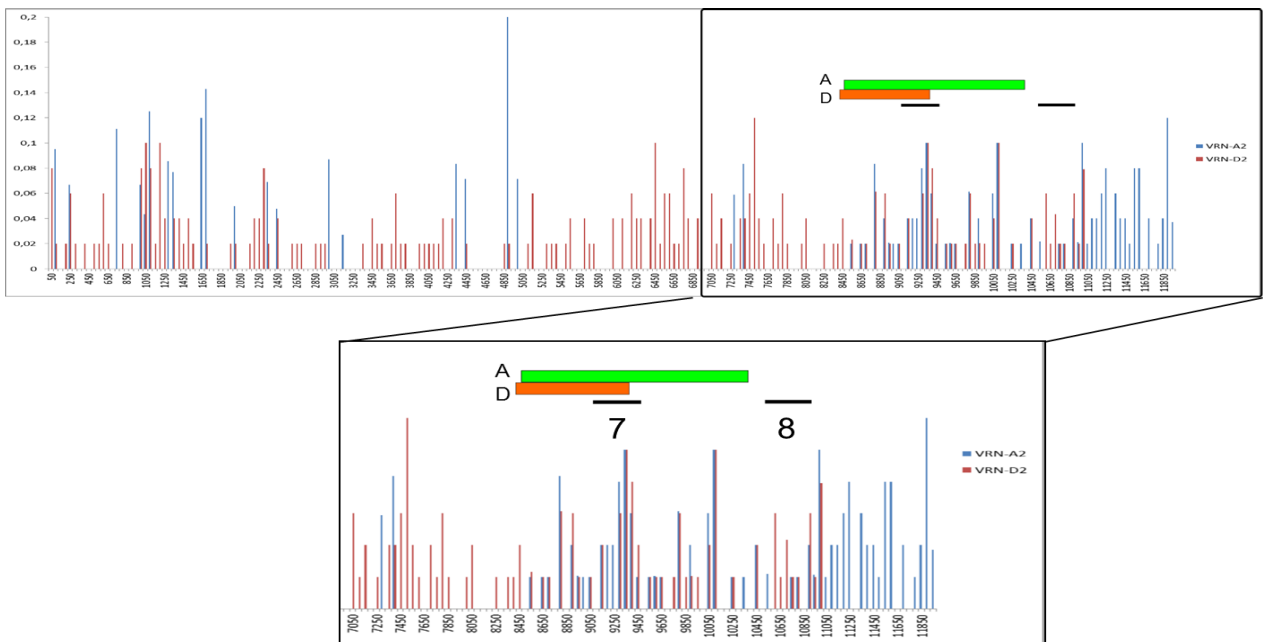


Figure 39A : Résumé des résultats de crible de méthylation, des profils CpG et des zones choisies pour l'étude des profils de méthylation

Les rectangles verts symbolisent la présence de méthylation, la couleur jaune l'absence et la couleur orange des résultats non concordant entre les deux méthodes de crible enzymatique.

Les traits noirs correspondent aux amplicons choisis et les numéros d'amplicons associés.

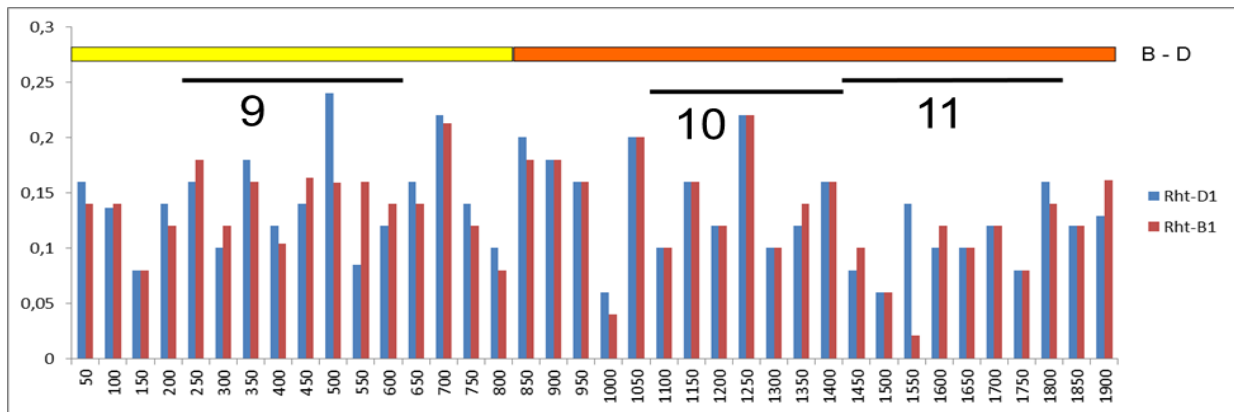
400-700bp. Un étudiant de Master 1 en Génétique et Physiologie a d'ailleurs développé un pipeline bioinformatique pour analyser les résultats de séquençage 454 après traitement au bisulfite (Stage de Mickaël Dumas, M1 AMDSV, été 2012). Toutefois, la technique Solexa est en train de supplanter la technique 454 de par son efficacité et Biogemma a décidé en 2013 de recentrer ces efforts de séquençage haut débit sur la technique Solexa/Illumina (Bennett, 2004) et d'abandonner la technique 454. Cette technique offre l'avantage d'effectuer un grand nombre de séquences par expérience (15 millions de séquences contre 400 000 pour le 454) et est moins sensible aux régions riches en homopolymères et donc sera moins sensible à la richesse en thymine des matrices après conversion au bisulfite. La taille des fragments séquencés est en constante augmentation, elle était de 35bp au départ mais avoisine maintenant les 400-500pb.

Le séquençage haut débit de nos 34 amplicons, devrait en outre nous donner accès aux copies homéologues comme ceci a été le cas lors de l'utilisation de la technique 454.

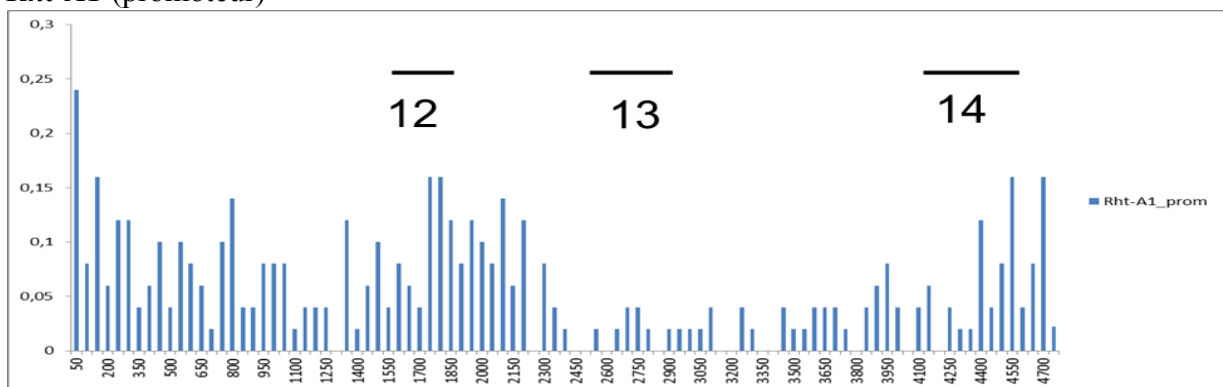
2.2.1.2. Stratégie d'amplification

La méthode de séquençage Solexa/Illumina nécessite le rajout d'adaptateurs distincts de part et d'autre de l'amplicon appelé adaptateur universel (UA) et adaptateur indexé (IA). Ce dernier contient dans sa séquence un index de 6 nucléotides spécifique à chaque IA. Dans le cadre de notre projet, un index donné sera choisi pour chaque génotype étudié soit au total 8 index distincts. Le rajout des adaptateurs se fait au cours de deux PCR imbriquées (voir Matériels & Méthodes). Lors de la seconde PCR afin d'obtenir la quantité de produit désirés, il est préférable d'augmenter la quantité de matrice de PCR1 plutôt que d'augmenter le nombre de cycles d'amplification et d'éviter l'effet de « clonalité ». Ainsi, après séquençage, la redondance d'une séquence sera due au fait qu'elle est abondante dans l'ADN utilisé comme matrice de départ et non à la reproduction en plusieurs exemplaires d'une séquence lors de la PCR2 suite à un trop grand nombre de cycles. Cette stratégie discutée avec Biogemma devrait limiter le biais d'amplification entre les différentes séquences amplifiées.

Rht-B1 et *Rht-D1*



Rht-A1 (promoteur)



Vrn-3 FT

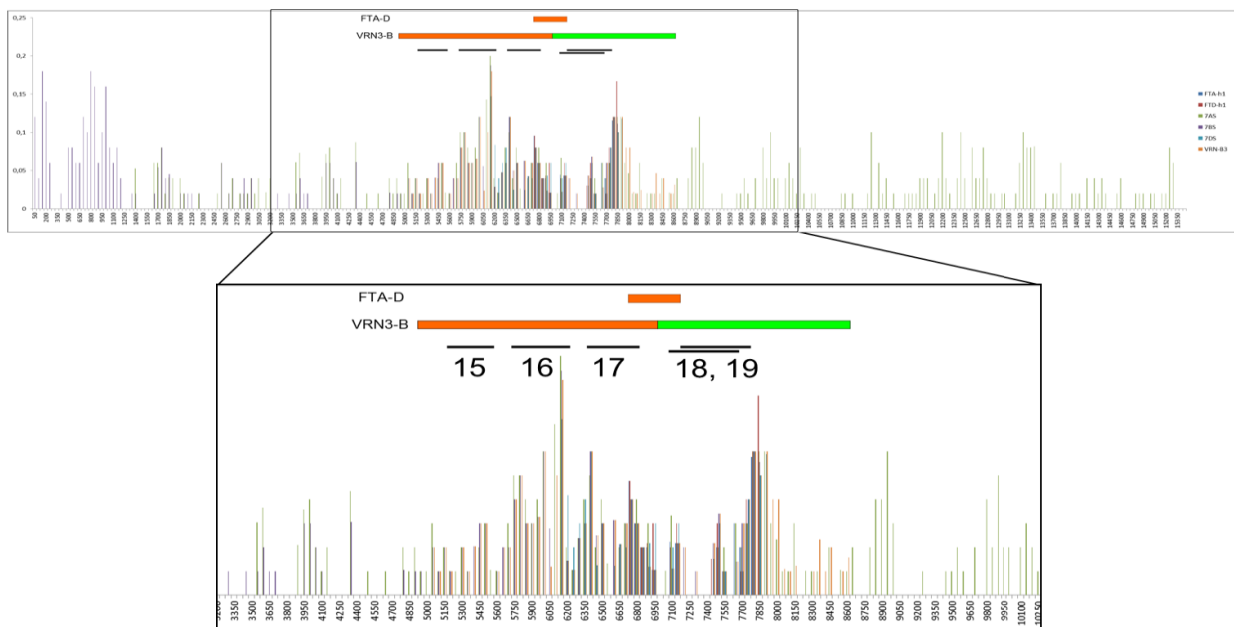


Figure 39B : Résumé des résultats de crible de méthylation, des profils CpG et des zones choisies pour l'étude des profils de méthylation

Les rectangles verts symbolisent la présence de méthylation, la couleur jaune l'absence et la couleur orange des résultats non concordant entre les deux méthodes de crible enzymatique.

Les traits noirs correspondent aux amplicons choisis et les numéros d'amplicons associés.

2.2.1.3. Mise au point des amplifications

La mise au point des PCR1 et PCR2 a été faite sur de l'ADN génomique de la variété *Chinese Spring* après traitement au bisulfite et en utilisant deux couples d'amorces déjà disponibles et spécifiques du gène *Vrn-1* : les amplicons *VRNI-1* (non méthylé) et *VRNI-3* (méthylé en contexte CpG) (Fig 40). Pour la PCR1, il n'est pas nécessaire de modifier les conditions de PCR utilisées précédemment avec des amorces sans adaptateurs. Le seul paramètre qu'il nous a fallu mettre au point est la quantité d'ADN issu de la PCR1 à utiliser pour la PCR2. J'ai également déterminé le nombre de cycles minimal à utiliser lors de la réamplification. Après ces mises au point, j'ai utilisé les 34 couples d'amplicons. Certains couples ne montrent aucune amplification ou une amplification non spécifique en PCR1 (Fig 41). Pour ces amplicons, plusieurs programmes PCR ont été testés sans pouvoir malheureusement résoudre ces problèmes d'amplification. La PCR2 (Fig 42) montre que 10 cycles d'amplification sont suffisants et qu'il vaut mieux augmenter la quantité de matrice PCR1 pour la PCR2. Certains des couples ne montrant pas d'amplification en PCR1 montrent une légère bande en PCR2. Une solution pourrait donc être d'augmenter la quantité de matrice en PCR1 pour les couples montrant ces profils (ces expériences sont actuellement en cours). L'ensemble des couples qui seront utilisés pour l'étude par conversion au bisulfite suivi d'un séquençage MiSeq sont résumés en Annexes.

2.2.2. Bilan des génotypes étudiés

Afin d'étudier nos régions d'intérêt, nous disposons

- de deux lignées aneuploïdes (2BS-1 et DT2BL) décrites au paragraphe 1.3. avec comme témoin la variété *Chinese Spring*.
- de quatre lignées RNAi de génération T2 descendant de la lignée T0 B0053038 produisant des siRNA (voir paragraphe 1.2.) et de leur témoin non transformé (lignée *NBI*).

A chaque génotype correspond un index soit au total 8 indexes différents.

2.2.3. Production des amplicons pour séquençage

Une fois la mise au point faite, l'étape suivante a été de produire les amplicons sélectionnés pour chacun des génotypes (réalisation en février 2014). Les ADN ont été extraits dans des conditions similaires aux analyses précédentes, il en fut de même pour les

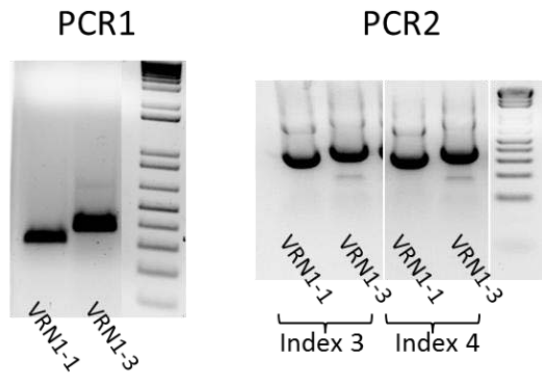


Figure 40 : Résultats de la mise au point des PCR1 et PCR2 avec les amplicons VRN1-1 et VRN1-3

La PCR2 a été faite avec deux amorces IA portant deux index différents : l'index 3 et l'index 4. La séquence de l'index n'influe pas sur l'efficacité de la PCR2.

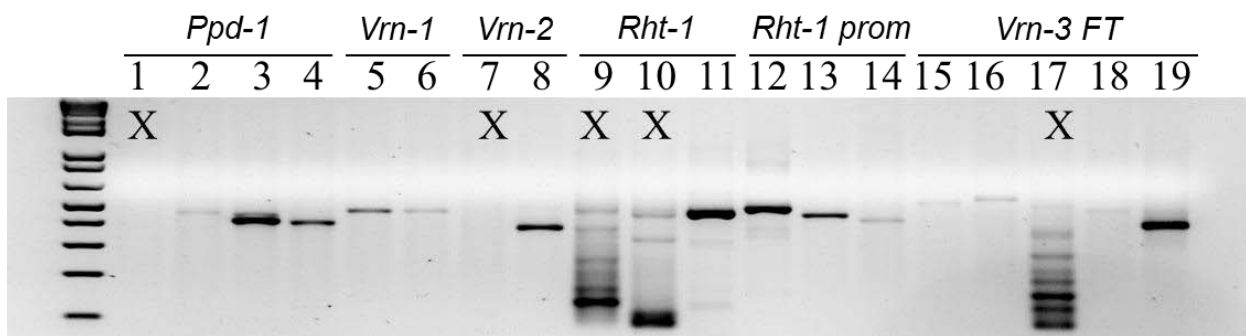


Figure 41 : Résultats de PCR1 pour les gènes cibles

Les croix X indiquent les amplicons qui ne montrent pas d'amplification ou une amplification non spécifique.

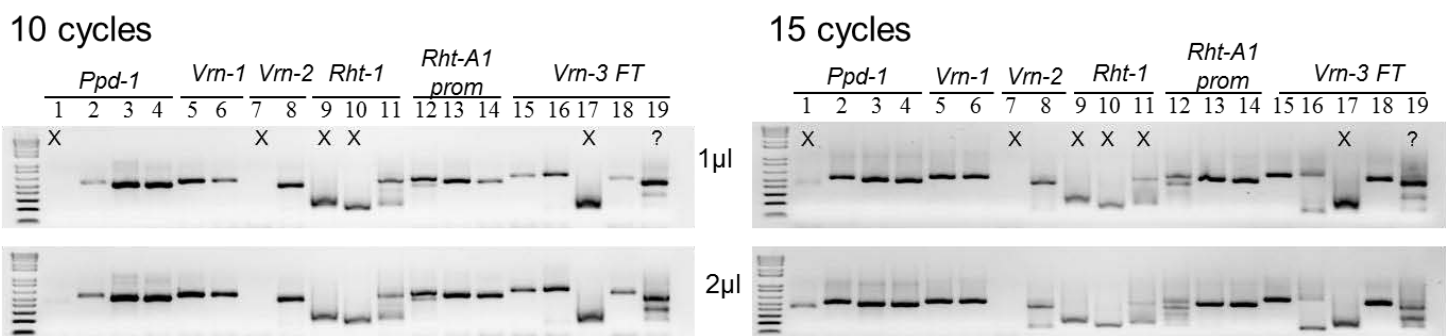


Figure 42: Résultats des PCR2 pour les gènes cibles

Deux quantités de PCR1 ont été utilisées comme matrice pour la PCR2, 1 et 2µl; deux durées de programme ont été testées, 10 et 15 cycles totales

conversions au bisulfite. Des problèmes d'amplification ont été observés, la présence importante de dimères d'amorces en PCR1, non visibles lors des mises au point. Cette première amplification de mauvaise qualité conduit à de très mauvaises amplifications en PCR2. Ces mauvaises amplifications ne sont pas retrouvées que pour les génotypes potentiellement modifiés sur leur épigénomes, mais également pour le génotype *Chinese Spring*, sur lequel ont été faites toutes les mises au point (Fig 43). Les premiers tests effectués montrent que la variable la plus limitante est la quantité d'ADNg utilisée lors de la conversion au bisulfite. Par manque de temps, il ne m'a pas été possible de produire les amplicons et d'effectuer le séquençage avant la rédaction de ce rapport. Nous espérons qu'ils seront disponibles lors de la soutenance de thèse.

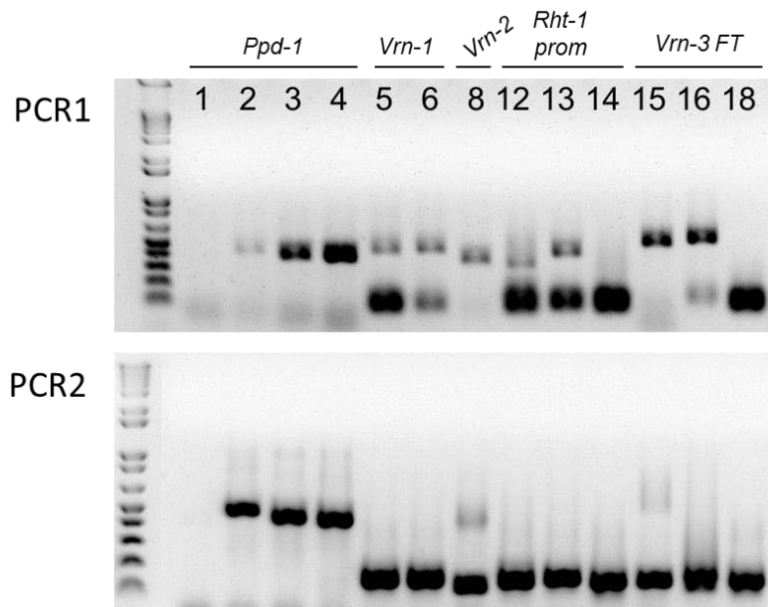


Figure 43 : Amplification PCR1 et PCR2 sur *Chinese Spring* pour la production des amplicons pour le séquençage MiSeq

Les amplicons 5, 6, 13, 14 et 18 montrent une forte bande de dimères d'amorces en PCR1 se se traduisant par une non amplification des amplicons voulus en PCR2.

Les amplicons 8 et 15 ne montrent pas de dimère d'amorces en PCR1, mais la bande de dimères d'amorces réamplifiés se retrouve en PCR2.

Discussion & perspectives

1. Quelles sont les principales conclusions de mon projet de thèse ?

De nombreuses connaissances sont disponibles chez *Arabidopsis*. Mon équipe d'accueil au GReD est spécialisée dans l'épigénétique et ceci nous a permis tout au long de ma thèse de pouvoir nous appuyer sur les méthodes et les connaissances du laboratoire afin de les appliquer chez une espèce cultivée, le blé.

1.1. Le gène *MET1* est très conservé au cours de l'évolution et présent chez le blé

Le démarrage de notre projet impliquait de rechercher le gène *MET1* chez le blé à partir d'orthologues connus chez d'autres espèces. Le gène *MET1* étant très conservé au cours de l'évolution, nous n'avons eu aucun mal à identifier les protéines et gènes *MET1* d'*Arabidopsis*, du riz, de l'orge ou encore de *Brachypodium* lors de nos recherches de similarité par BLAST. Les protéines de la famille *MET1/DNMT1* sont d'ailleurs conservées chez de nombreuses espèces comme l'Homme, l'abeille, certains champignon et les plantes (Jeltsch, 2010) avec dans tous les cas un rôle fonctionnel dédié à la maintenance de la méthylation de l'ADN. Non seulement la méthylation est présente chez de nombreuses espèces mais les profils de méthylation au niveau du corps des gènes semblent conservés entre espèces. Ainsi, la comparaison des profils de méthylation dans le corps des gènes chez le riz, *Arabidopsis* et *Brachypodium* montrent des profils très similaires entre les gènes orthologues malgré une divergence de plusieurs millions d'années (Takuno and Gaut, 2013). Chez le blé tendre, il nous reste toutefois à valider la fonction des neuf copies de *TaMET-1*, et entre autre de démontrer que les copies du groupe 2, qui sont les plus exprimées, assurent l'essentiel de la fonction de maintien de la méthylation des CpG. La lignée RNAi établie au cours de la première vague de transgénése et qui produit des siRNA ou les futures lignées de la seconde vague devraient représenter de bons outils pour réaliser cette validation.

1.2. Le blé compte neuf copies de *TaMET-1*

Nous avons été un peu surpris de trouver 9 copies du gène *MET1* chez le blé alors qu'il n'en existe qu'une seule chez *Arabidopsis*. Ceci nous a conduit à nous intéresser aux mécanismes à l'origine des duplications de gènes.

1.2.1. Elimination des paralogues après une Whole Genome Duplication (WGD)

Après une duplication, qu'elle soit à l'échelle du génome ou d'un segment plus restreint, l'un des deux paralogues est le plus souvent éliminé au cours de l'évolution. Par exemple les premières études réalisées chez la levure montrent que seuls 12% des gènes dupliqués sont conservés après une WGD survenue il y a environ 150MYa (Kellis et al., 2004). Chez le maïs, qui aurait subi une WGD il y a seulement 5MYa (Salse et al., 2008a), il semblerait que pour 50% des gènes dupliqués, un des deux paralogues ait déjà été éliminé (Lai et al., 2004). La perte des gènes serait donc massive et se ferait très rapidement après la duplication. Un premier constat est que chez plusieurs espèces ayant connues des WGD comme le riz, la carotte, *Brassica* et le blé, les paralogues de *MET1* semblent ne pas être éliminés. Ainsi, le riz possède deux paralogues du gène *MET1* (*OsMET1a* et *OsMET1b*) issus de la WGD commune à toutes les Graminées. Dans ce cas, et malgré le fait que cette duplication remonte à plus de 50-70MYa, les deux copies du gène ont été conservées et sont toutes deux fonctionnelles (Teerawanichpan et al., 2004). Chez les *Brassica* (Fujimoto et al., 2006) et la carotte (Bernacchia et al., 1998), deux paralogues sont également présents et sont exprimés dans des tissus différents. *MET1* n'est pas le seul cas où les paralogues sont conservés : d'autres gènes comptent un nombre de copies bien plus important qui sont maintenues après duplication. A titre d'exemple, la famille de gènes codant pour des facteurs de transcription à domaine homéotique leucine zipper (*HD-Zip*) chez le soja est constituée de 88 gènes issus de multiple duplication à l'échelle du génome ou de segments chromosomiques (Chen et al., 2014). Globalement tous ces gènes sont exprimés mais montrent des différences de profils en fonction des tissus ou des conditions de stress (Chen et al., 2014). Il y aurait donc un intérêt pour certains organismes à conserver les différents paralogues après des duplications. Le paragraphe suivant propose une explication évolutive à cette observation.

1.2.2. Devenir des paralogues si ils ne sont pas éliminés

Lorsque les paralogues échappent à l'élimination massive comme celle observée chez la levure, plusieurs scénarios sont possibles. Les deux versions du gène peuvent (i) conserver leur fonction initiale, (ii) acquérir une complémentarité pour assurer la fonction, (iii) l'un des paralogues peut diverger vers une nouvelle fonction ou (iv) devenir non fonctionnel. Ces différents cas de figures sont décrits dans la littérature sous les termes de redondance, sub-fonctionnalisation, néo-fonctionnalisation et pseudogénéisation (Lawton-Rauh, 2003). Il semble que pour *TaMET-1* plusieurs cas de figure aient eu lieu. Comme nous l'avons déjà expliqué, l'ensemble des paralogues issus des deux duplications successives ont été conservés pour *TaMET-1*. Concernant le groupe 5, les copies homéologues présentes sur les chromosomes 5A et 5D sont devenues des pseudogènes puisque ces deux gènes présentent des délétions dans leur séquence codante et ne montrent aucune expression. La copie *TaMET-5B1* pourrait avoir suivie le scénario d'une néo-fonctionnalisation puisque une expression est observée dans le grain au stade 350 degrés jour après floraison. L'expression grain spécifique du gène *TaMET-5B1* est d'ailleurs plus importante que pour les copies du groupe 2. Toutefois, le taux d'évolution de la copie 5B (et notamment l'accumulation de mutations non-synonymes) suggère qu'elle serait également engagée vers la pseudogénéisation. Enfin, pour les copies du groupe 7, il est difficile de choisir entre sub-fonctionnalisation et néo-fonctionnalisation.

Une hypothèse serait donc que la conservation des paralogues de *MET1* pourrait contribuer à la diversification du profil d'expression de ce gène.

1.3. Les gènes *TaMET-1* du groupe 2 sont exprimés dans les tissus actifs pour la réplication

Une des caractéristique des gènes de la famille *MET1* est de maintenir la méthylation des cytosines, c'est-à-dire de méthyler le brin néo-synthétisé lors de la réplication (Pradhan et al., 2008). Il est donc attendu une forte expression de *MET1* dans les tissus à forte activité de réplication. Dans notre étude, les profils d'expression ont été faits sur quinze échantillons au cours du développement du blé. Cet ensemble de tissus bien que constituant déjà un bel échantillonnage n'est pas exhaustif, il est donc difficile d'associer l'expression de nos gènes *TaMET-1* avec des tissus à forte prolifération comme les méristèmes. L'expression retrouvée dans les échantillons des tiges, qui incluent également le méristème apical, et dans les épis en

cours de formation vont dans ce sens. Concernant la copie *TaMET-5B1*, il est possible qu'elle s'exprime dans d'autres tissus qui n'ont pas été étudiés ou dont l'expression est « noyée » dans un prélèvement qui contiendrait un trop grand nombre de tissus différents. De même, il est possible que l'expression des copies du groupe 7 soit plus importante que les copies du groupe 2 dans d'autres tissus non étudiés. A titre d'exemple, les tissus racinaires, qui sont un second tissu à forte division avec la présence de l'apex racinaire (méristème racinaire), n'ont pas été intégrés à l'analyse. Une étude en RT-qPCR pourrait être réalisée dans le futur. Cette étude bénéficierait des amorces PCR déjà définies notamment permettant de distinguer les trois copies homéologues du groupe 2.

1.4. Il existerait une régulation transcriptionnelle de la famille *MET1* par la méthylation

Bien que l'ensemble des paralogues soit encore présents, nous avons montré que certaines des copies de *TaMET-1* ne s'expriment pas. Nous avons pu démontrer par une analyse de conversion au bisulfite que les promoteurs des groupes 5 et 7 présentaient une méthylation plus ou moins importante de leurs promoteurs. Cette méthylation pourrait être responsable de la répression transcriptionnelle des gènes *TaMET-1* des groupes 5 et 7. Nous n'en avons pour l'instant pas la démonstration formelle.

Nous avons tenté au cours de mon projet de thèse de traiter des plantes de la variété *Chinese Spring* avec un agent déméthylant ou en tout cas bloquant le maintien de la méthylation : la 5-Azacytidine (5-AZA) (Amado et al., 1997). Nos expériences n'ont pas été concluantes car il s'est avéré difficile de définir les conditions précises d'utilisation de la 5-AZA (dose, méthode d'application, stade de croissance des plantes...). Les plantes traitées bien que montrant des défauts de croissance, n'ont montré aucune modification du profil de méthylation de notre amplicon test VRN1-3 riche en cytosines méthylées dans le contexte CpG (données non montrées). Ce type d'expérience serait très intéressante à reproduire dans le futur afin de démontrer une possible levée de la répression transcriptionnelle des gènes du groupe 7. Une autre alternative serait de réaliser des expériences d'expression transitoires afin de démontrer que les régions promotrices des gènes du groupe 7 (et éventuellement du groupe 5) sont toujours fonctionnelles.

Une étude récemment réalisée chez les mammifères a établi une corrélation négative entre la présence de méthylation au niveau des promoteurs des gènes dupliqués et le taux d'évolution

(Keller and Yi, 2014b). En effet les auteurs ont montré que les paires de gènes résultants des plus anciennes duplications sont moins méthylées au niveau de leur promoteur et qu'il peut exister de grandes différences dans les profils de méthylation pour les deux duplicats. A l'inverse, les gènes dupliqués plus récemment présentent un taux de méthylation similaire au niveau de leur promoteur, méthylation qui dans ce cas est importante et inversement corrélée avec le niveau d'expression (Keller and Yi, 2014b). Ces observations sont cohérentes avec les profils des régions promotrices des trois groupes de copies de *TaMET-1*. La duplication la plus ancienne est la duplication w2-w5. Nous n'avons malheureusement pas formellement démontré la faible méthylation des promoteurs du groupe 2 (quelques clones issus de conversion au bisulfite indique une non méthylation des CpG) mais la faible représentation en CpG laisse penser que le niveau de méthylation sera faible. Inversement, la méthylation des régions promotrices des groupes 5 et 7, duplication la plus récente, est plus importante, plutôt homogène entre les deux régions promotrices et associée avec une faible expression, alors que les copies du groupes 2 sont les plus exprimées. Nos observations chez le blé sont donc cohérentes avec celles de Keller et Yi (2014).

1.5. Existe-t-il une autorégulation dans la famille MET1 chez le blé ?

Un des résultats surprenant est l'enrichissement du promoteur de la copie *TaMET-7D1* en dinucléotides CpG. Les CpG étant ciblés par MET1, il est envisageable que les protéines MET1 codées par les copies du groupe 2 assurent leur méthylation. Les copies du groupe 2 échapperaient à une telle boucle de régulation car leurs promoteurs comportent peu de CpG. S'agit-il d'une boucle d'autorégulation au sein de la famille *MET1* ? Le profil de méthylation de *MET1* chez *Arabidopsis* est connu (<http://neomorph.salk.edu/epigenome/epigenome.html>) mais ne montre pas d'enrichissement de la méthylation en CpG dans sa région promotrice. Il n'y a pas de données disponibles chez la souris, l'Homme ou le riz, il est donc difficile de conclure si il y a un lien entre enrichissement en CpG des régions promotrices, méthylation et régulation transcriptionnelle des copies de *MET1*.

Le rôle d'une telle boucle, si elle existe, reste difficile à comprendre. Une forte expression du groupe 2 régulerait négativement les copies du groupe 7 : celles-ci ne seraient donc pas exprimées dans les tissus exprimant fortement le groupe 2. Ce mécanisme pourrait participer à la sub-fonctionnalisation ou néo-fonctionnalisation du groupe 7 dont l'expression selon un profil différent permettrait d'assurer la fonction de *MET1* en toute circonstance.

1.6. La perte de fonction de MET1 conduirait à une létalité et/ou une stérilité

1.6.1. Constat suite à mon projet de thèse

Nous avons été confrontés à un problème important au cours de mon projet : la surmortalité de nos lignées RNAi. Cette létalité est sans doute une réalité, en effet chez le riz une récente publication indique qu'il est impossible d'obtenir des mutants homozygotes pour *osmet1b*, copie la plus exprimée, alors que c'est le cas pour *osmet1a*. De plus, la perte de fonction d'*OsMET1b* n'est pas sauvée par *OsMET1a* (Yamauchi et al., 2014). Le même phénomène pourrait avoir lieu chez le blé, et les copies du groupe 2 (orthologues d'*OsMET1b*) seraient essentielles pour la viabilité de la plante. D'après les expériences menées chez le riz, les copies du groupe 7 (orthologues d'*OsMET1a*) ne seraient dans ce cas pas capables d'assurer la survie des plantes dans le cas d'une perte de fonction des copies du groupe 2, comme cela peut être envisagé dans nos lignées RNAi. Notons également que notre échec pour identifier un mutant TILLING de *MET1* pourrait s'expliquer par le fait que nous n'avons recherché que des mutations nulles (codons stop) en ciblant préférentiellement les copies du groupe 2. A posteriori, ces mutants nuls de *MET1* pourraient s'avérer létaux et donc impossible à cribler. Il faut toutefois rappeler que notre méthode de crible n'était pas optimale. Par exemple nous n'avons recherché que des codons stop représentant seulement 5% des mutations dans une population de TILLING. D'autre part, les amorces utilisées pour notre crible par la méthode KASPAR ont été définies à un stade où nous n'avions pas connaissance de l'existence des neuf copies de *MET1* (ou ne tenait pas compte des similitudes entre les copies homéologues). Nos amorces amplifiaient donc peut-être plusieurs copies simultanément et n'auraient pas permis de détecter les mutants.

1.6.2. Comment améliorer notre démarche ?

Il nous semble que la stratégie RNAi reste la plus prometteuse compte tenu de l'existence de plusieurs copies exprimées des gènes *TaMET-1*. Chez *Arabidopsis* la stratégie RNAi contre *MET1* a été réalisée avec le promoteur de *DEMETER* (*DME*), qui s'exprime au de façon très précoce dans le développement des graines et plantules (Kim et al., 2008). Cette fenêtre d'expression très ciblée suffit à diminuer durablement la méthylation au cours de la vie végétative. Notons que le gène *DME* est connu (Wen et al., 2012) chez le blé et que son profil d'expression connu chez l'orge semble assez proche de celui décrit chez *Arabidopsis*

(Kapazoglou et al., 2013). Il serait donc envisageable de vérifier le profil d'expression de *DME* chez le blé en utilisant les données du RNA-seq, et de réaliser une stratégie RNAi sous contrôle du promoteur *DME* comme chez *Arabidopsis*. Chez l'orge d'autres systèmes inductibles ont été utilisés comme les promoteurs inductibles par le froid issus des gènes *WRKY71* du riz et *Cor39* du blé dur (Kovalchuk et al., 2013). Notons également que le promoteur du gène *Hvhsp17* de l'orge inductible par choc thermique a aussi été utilisé avec succès chez le blé (Freeman et al., 2011). De tels promoteurs avec une expression tissu spécifique ou inductibles par le froid ou la chaleur pourraient être utilisés pour construire une troisième vague de RNAi dirigés contre *MET1*.

1.7. Questions futures

Beaucoup de questions restent en suspend à la fin de mon projet. Certaines parties du travail sont à finaliser comme celle concernant les gènes candidats ou la caractérisation de la seconde vague de lignées RNAi. D'autres pistes de recherche pourraient également être envisagées dans le futur comme rechercher des partenaires de MET1 participant à la méthylation des CpG, cibler d'autres gènes impliqués dans le maintien des méthylations comme par exemple *DDMI*, cibler d'autres voies impliquées dans la méthylation *de novo* telle que le RdDM ou impliquée dans la déméthylation active.

2. Mon projet de thèse aura-t-il des applications futures chez Biogemma ?

2.1. Méthode d'évaluation des lignées RNAi via l'efficacité de production de siRNA

Mon travail de thèse a permis entre autre de mettre en place la validation des construits RNAi par recherche des sRNA. Lors de la validation des lignées transgéniques, le nombre de copies insérées ainsi que l'intégrité du transgène sont analysés. Pour valider les lignées RNAi nous aurions pu regarder l'expression de ce transgène, ce qui aurait nécessité beaucoup de travail de mise au point et ne nous aurait pas assuré de l'efficacité de la construction. La recherche de sRNA permet de s'assurer que le transgène est fonctionnel et permet la production des intermédiaires qui assurent la mise sous silence de la cible. Dans notre cas cela

s'est avéré particulièrement utile : si la sélection ne s'était faite que sur l'intégrité des construits, aucune descendance n'aurait été conservée à partir du stade T1. De plus, deux plantes T0 présentaient un construit tronqué, mais une seule a donné une descendance produisant des sRNA. Cela a permis de sélectionner une lignée où la production de sRNA connus pour induire le silencing a été démontrée.

L'intégration de la recherche de sRNA dans la procédure de caractérisation des lignées transgéniques produites à Biogemma pourrait se révéler peu efficace si elle est réalisée par Northern Blot. Il serait toutefois possible d'envisager une extraction des ARN totaux sur un grand nombre d'évènements T0, d'isoler les sRNAs et de les séquencer. Biogemma possède en effet une plateforme de séquençage haut débit MiSeq qui permettrait une caractérisation rapide des sRNA. L'analyse des résultats permettrait de vérifier la production de sRNA correspondant à la cible, mais également de mettre en évidence d'éventuelles modifications dans les profils d'expression, dont celui de la cible.

Chez l'Homme, il existe un système de validation de l'efficacité des construits RNAi qui se base sur la luciférase (RNAi validation system, OriGene). La séquence ciblant le gène d'intérêt est insérée dans un vecteur à la suite du gène de la luciférase (LUC). Un ARN chimérique est produit à partir du transcrit LUC et de la séquence insérée, qui formera la structure en double brin. Si la machinerie du silencing prend en charge l'ARNdb, l'ensemble de l'ARN chimérique sera détruit. Dans le cas d'un RNAi efficace, il n'y aura pas production de la protéine LUC et donc pas de lumière émise. Un système similaire pourrait être utilisé chez les plantes, en utilisant le gène LUC ou un autre gène menant à la production d'un composé coloré par exemple.

2.2. Méthode d'étude du méthylome chez le blé par conversion au Bisulfite et séquençage

Au cours de ce travail, j'ai eu l'occasion de mettre au point le crible des cibles à partir d'enzymes sensibles à la méthylation suivi d'une amplification. La seule limitation à l'utilisation du couple HpaII/MspI est la présence du site de coupure dans l'amplicon, ou l'utilisation d'amorces spécifiques à chaque génome homéologue, si le site de coupure n'est pas présent pour tous les génomes. Les résultats que nous avons obtenus en utilisant l'enzyme McrBC ont été moins concluants. L'une de ces raisons est peut-être un problème d'accessibilité du génome par l'enzyme.

J'ai également mis en place un protocole standard de conversion au bisulfite à partir d'ADN génomique de blé (Annexes). Au départ du projet de thèse, il était prévu de réaliser un séquençage de type 454 après la conversion au bisulfite, seule technologie de séquençage haut débit disponible à Biogemma. Les premières mises au point ont donc été réalisées avec cette technologie. Nous avons également participé au stage de Mickaël Dumas qui a développé un outil informatique pour le BS-Seq. Mickaël a validé son outil en utilisant les amplicons VRN1. Mais au cours de ces trois ans, Biogemma a acquis une nouvelle technique de séquençage haut débit permettant de générer un plus grand nombre de séquences, de meilleures qualités tout en conservant des longueurs de séquence similaires, il s'agit de la technologie MiSeq d'Illumina. Le séquençage de type 454 a donc été abandonné et il a fallu s'adapter à ce changement. Ce changement est en partie responsable du fait qu'actuellement les résultats de séquençage sur les lignées d'études ne sont pas disponibles. Les derniers résultats obtenus montrent que la production des amplicons qui nécessitent des amorces contenant les adaptateurs nécessaires au séquençage pose quelques problèmes. Les premiers tests réalisés vont dans le sens de l'importance de la quantité d'ADN convertis, ce qui devrait être réglé rapidement.

2.3. Modifier le méthylome du blé en ciblant *DDMI*

L'objectif initial de ma thèse était de travailler sur les gènes *MET1* et *DDMI*, connus chez *Arabidopsis* pour leur implication dans le maintien de la méthylation de l'ADN. Au cours de l'étape d'identification des séquences de ces gènes chez le blé tendre, il n'a pas été possible d'obtenir les séquences de *DDMI*. La suite du projet s'est donc focalisée sur *MET1*. Aujourd'hui, grâce à l'apport des données IWGSC, il serait possible d'obtenir les séquences de *DDMI*. Un tel projet bénéficierait des contigs incomplets générés lors de nos expériences de Sequence Capture, des résultats de cartographie déjà obtenus dans les populations ReR et Bal, et des séquences chez des espèces proches notamment les deux espèces de blé diploïde dont la séquence a été publiée en 2013. De la même manière que pour *MET1*, il serait possible de créer des lignées RNAi ciblant *DDMI* ou de cribler différentes populations de mutants. Chez *Arabidopsis*, les mutants *ddm1* montrent une diminution de la méthylation dans tous les contextes de séquences. Le gène *DDMI* est un acteur important dans la répression des éléments transposables, en assurant l'accessibilité aux méthyltransférases des régions hétérochromatiques, il agit de ce fait également sur la régulation de l'expression des gènes

hétérochromatique (Zemach et al., 2013). Au vue de la proportion des éléments transposables dans le génome du blé, l'effet d'une mutation *ddm1* pourrait avoir des effets très important, beaucoup plus important que la mise sous silence de *met1*. C'est pourquoi, cibler le gène *DDMI* reste intéressant mais en privilégiant une approche RNAi, avec éventuellement utilisation d'un promoteur inductible. Une telle lignée RNAi pourrait également être utilisée pour faire de la mutagenèse par insertion d'éléments transposables chez le blé.

2.4. Effets possibles sur la dérèpression des éléments transposables dans les mutants *met1* et *ddm1*

Dans les mutants *met1* et *ddm1* d'*Arabidopsis*, la transcription ainsi que les mouvements de certains élément transposable sont observés, comme par exemple l'élément EVD (Mirouze et al., 2009) ou les éléments de type CACTA (Reinders et al., 2009). Des phénomènes similaires ayant été observés chez le riz (Yamauchi et al., 2014), la répression de *MET1* chez le blé pourrait avoir des effets importants sur les éléments transposables qui représentent un peu plus de 80% du génome du blé et en particulier sur les éléments CACTA qui représentent à eux seuls 14% des éléments transposables du génome du blé (Choulet et al., 2010)

Tout comme chez *Arabidopsis*, les éléments transposables sont probablement fortement méthylés et la perte de la méthylation aura donc un impact très fort sur cette fraction du génome. La dérèpression des éléments transposables chez le blé, ne se fera pas à la même échelle que celle observée chez *Arabidopsis*, il est donc difficile de prévoir l'effet sur la stabilité du génome. La dérèpression de ces éléments pourrait être source de nouveaux phénotypes d'intérêt ou source de mutations, reste à savoir si de telles plantes seront viables.

2.5. Création d'EpiRIL *met1*

Une population EpiRILs pourrait être créée à partir d'un mutant *met1* et d'un sauvage. Les lignées RNAi ont un effet possible sur les profils de méthylation, et en tant que mutant pourrait être utilisée directement comme lignée mutée. En se basant sur la manière dont ont été établie les EpiRILs chez *Arabidopsis*, il faudrait 8 générations pour obtenir des lignées stabilisées soit près de 10 ans d'expérimentation.

Dans le cas ou des mutants TILLING ou d'irradiation seraient identifiés, ils seraient également envisageable de les utiliser comme mutant *met1* pour le croisement initial. Pour

être sûr d'avoir un effet, il sera sans doute nécessaire d'avoir des mutants pour les trois copies homéologues du groupe 2. Ceci supposerait de les regrouper dans une même lignée par croisement puis d'établir une lignée homozygote. Au préalable, sachant que les lignées de TILLING contiennent un grand nombre de mutations, il serait peut être nécessaire d'effectuer plusieurs génération de backcross afin d'éliminer la présence d'autres mutations éventuelles avant l'utilisation dans la création d'EpiRIL à proprement parlé. Ce sont donc des approches de longues durées.

A partir de ces EpiRILs : il sera possible d'identifier des caractères/phénotypes montrant des variations stables, liées à une modification du profil de méthylation. Ainsi, ces polymorphismes de méthylation identifiés pourraient être utilisés de la même manière que des polymorphismes de séquences.

2.6. Utilisation de lignées RNAi pour changer les profils de recombinaison

Les modifications des profils de recombinaison liées à des différences de méthylation et en l'absence de polymorphisme nucléotidique ont été étudiées chez *Arabidopsis*. Pour ce faire, des lignées EpiRILs issues d'un mutant *ddm1* ont été utilisées pour identifier des régions différenciellement méthylées entre le sauvage et l'épi- mutant. De telles modifications doivent être stables au cours des méioses. Les analyses ont montré que la perte de méthylation au niveau des séquences répétées s'associait avec une suppression totale des crossing overs (COs) au niveau des régions péricentromériques, suppression qui s'étend sur 3Mb au niveau des régions flanquantes. L'analyse du nombre de COs et de leur répartition a montré qu'il y avait une augmentation de la taille des distances génétiques au niveau des bras chromosomiques chez les EpiRILs (Colomé-Tatché et al., 2012).

Chez *Arabidopsis*, la perte de la méthylation favorise les recombinaisons au niveau des bras chromosomiques, qui correspondent aux régions géniques et donc aux régions faiblement méthylées. Chez les plantes sauvages, la recombinaison augmente fortement aux frontières entre les régions péricentromériques et les régions riches en gènes : cette caractéristique chez la plante sauvage semble fortement diminuée dans les EpiRILs (Colomé-Tatché et al., 2012). Comment ces résultats vont se transposer chez le blé ? Les régions péricentromériques chez *Arabidopsis* correspondent aux éléments répétés du génome ; chez le blé ces mêmes éléments se retrouvent tout le long des bras chromosomiques, alors que les gènes sont majoritairement

organisés en cluster. Chez le blé, les COs sont plus fréquents au niveau des télomères (région les plus riches en gènes) (Saintenac et *al.*, 2009), mais des variations à une échelle plus faible sont également notées qui dépendent du pourcentage de séquences codantes (Saintenac et *al.*, 2011b). Observera-t-on une augmentation des COs dans les régions plus riches en gènes dans des lignées modifiées pour leur profil de méthylation? En extrapolant ce qui est connu chez *Arabidopsis*, une augmentation des COs serait attendue au niveau des télomères. Un phénomène similaire pourrait être attendu au niveau des régions riches en gènes le long des bras chromosomiques. Toutefois, l'influence des ETs pourrait avoir un effet déterminant puisque selon les observations faites chez *Arabidopsis* se sont les limites région génique – région répétée qui seront modifiées. Ces régions sont beaucoup plus fréquentes chez le blé.

3. Conclusion

L'épigénétique est une discipline nouvelle et mon projet aura je l'espère contribué à mieux comprendre la fonction d'un acteur clef : la méthylation de l'ADN et le gène *MET1*. Il y aura beaucoup d'autres sujets qui pourront maintenant être abordés grâce aux nouvelles séquences IWGSC. Par exemple des expériences de CHIP-Seq permettraient de savoir si le code histone ainsi que la répartition des marques épigénétiques sont similaires à ce qui est décrit chez *Arabidopsis*. Il sera sans doute compliqué de définir la notion d'hétérochromatine chez le blé car chez cet organisme il y a plus de séquences répétées que chez *Arabidopsis* et les gènes qui doivent s'exprimer seront sans doute fortement influencés par celles-ci qui sont le plus souvent fortement méthylées.

Matériels & Méthodes

N° amorces	Nom amorce	séquence 5'3'	
Couple 1	met1_GPIC_F1	TTGATCAACAGAATATGGAGTG	
	met1_GPIC_R1	CACCATTACCAAGACCCAG	
Couples 2	met1_TA83026_F1	GCTTTGAAAAAGAACAAGGG	
	met1_TA83026_R1	GACCGCCATTTATGAATTC	
Couple 3	met1_TA83026_F2	GGCATGAACAGATTCAACC	Amorces ciblant les exons
	met1_TA83026_R2	TTACTTGCACCATTTCCAC	
Couple 4	met1_TA83026_F3	TTAGAGATACAATTGGGGATTG	
	met1_TA83026_R3	AACTGGTAGCTGTCAGGAAAG	
Couple 5	met1_TA83026_F4	GCAAGCACAGGCAGATTG	
	met1_TA83026_R4	CACACAAAAGACTTCTATCCCC	
Couple 6	met1 c6_F	ATGGAATATTTATGTAGAAGCCTG	Amorces pour compléter les régions manquantes
	met1 c6_R	AATGTCAAGAGTTGCAATACG	
Couple 7	met1 c7_F	TGAAATTGTTAATTTATGTGCTG	
	met1 c7_R	AAAATACTCCGCAAAGAGAG	
Couple 8	met1 c8_F	TGTTGTAAGGCATGGTAAATAATC	
	met1 c8_R	GAGAACTTGAATGCTGGTG	

Tableau 9 : Amorces utilisées pour l'amplification de *MET1* sur *T.monococcum*

La première partie du tableau donne les amorces ciblant les parties exoniques. La seconde partie du tableau donne les amorces qui ont permis de compléter les régions introniques manquantes.

Les données du matériel et méthodes sont communes aux chapitres 1 et 2.

1. Identification des ESTs spécifique à *MET1* et *DDM1* chez le blé.

A partir des séquences de *MET1* connues chez *Arabidopsis* (At5g49160), chez le riz (Os07g08500 & Os03g58400) et chez *Brachypodium* (Bradi1g55287 & Bradi1g05380) une recherche de similarité a été réalisé en utilisant le logiciel BLASTn (paramètres par défauts) contre l'ensemble des données de transcrits de blé.

La même recherche a été faite de la même manière, à partir des séquences de *DDM1* d'*Arabidopsis* (At5g66750), du riz (Os09g27060 et Os03g51230) et de *Brachypodium* (bradi1g10360).

2. Amplification de *MET1* chez *T. monoccocum*

La recherche des séquences génomiques correspondantes à *MET1* sur la base regroupant les données de séquençage de *Chinese Spring 5X* a été faite à partir des ESTs TA83026_4565 et GPIC:S:720428. Les contigs ESTs ont été alignés avec les séquences génomiques de blé et les séquences des gènes *Brachypodium* en utilisant l'outil SIM4 (Florea et al., 1998). Le tableau 9 reprend l'ensemble des amorces utilisées pour l'amplification de *MET1* chez *T.monoccocum*. Les produits PCR ont ensuite été séquencés auprès de GATC Biotech.

Les PCRs ont été réalisées dans les conditions suivantes :

Mix PCR	Concentration initiale	Volume (µl)
Green GoTaq® Reaction Buffer	5X	4
MgCl ₂ (Promega)	25mM	1,5
dNTPs	2,5mM chaque	1
Amorce sens	10µM	1
Amorce anti sens	10µM	1
Matrice ADNg	<200ng	
H ₂ O		qsp 20µl

Programme PCR :

95°C	2min	
95°C	30 sec	x 30 cycles
53°C	30 sec	
72°C	1 min	
72°C	5 min	
15°C	pause	

3. Crible des populations TILLING

3.1.1. Identification des codons stop potentiels

Pour la première vague de crible, les 3 séquences de *MET1* résultant du premier Sequence Capture (c1, c2 et c3 ; Fig 15) et la séquence de *MET1* amplifié sur *T.monococcum* ont été utilisées.

Dans un premier temps la séquence met1_c2, la plus longue des quatre séquences, a été alignée avec la séquence codante et la séquence protéique de *Brachypodium* la plus proche (bradi1g55287) en utilisant respectivement les outils SIM4 et GeneWise (EMBL-EBI : <http://www.ebi.ac.uk/Tools/psa/>). Les jonctions intron-exon ont été délimitées à partir de cet alignement pour produire un transcrit. Le cadre de lecture a été vérifié en alignant la traduction en séquence protéique de ce transcrit avec la protéine de *Brachypodium*.

A partir de l'alignement du transcrit de *MET1* du blé et de la protéine correspondante fait avec GeneWise, recherche des codons qui deviendront potentiellement des codons stop après un traitement à l'EMS : {CAA->TAA}, {CAG/TGG->TAG}, {CGA/TGG->TGA}.

Les codons identifiés ont été reporté sur l'alignement fait avec les 4 séquences de *MET1* du blé listées plus haut afin de déterminer lesquels sont conservés entre les 4 séquences.

Pour la seconde vague de crible seule la séquence du transcrit de la copie *TaMET-2B1* a été utilisée pour la recherche des codons stop potentiels, comme expliqué plus tôt.

3.2. Design des amorces utilisées pour le crible des populations TILLING

Pour les codons sélectionnés, le contexte de séquence a été récupéré, soit la base d'intérêt et 100pb avant et après cette base.

A partir de ces séquences, l'outil PrimerPicker (Kbioscience) définit les amorces nécessaires pour le génotypage: il fournit 2 amorces A1 et A2 qui ne diffèrent que par le dernier nucléotide et 2 amorces C1 et C2. Pour une amplification il est utilisé un mélange des amorces A1 et A2 avec C1 (ou C2 quand C1 ne fonctionne pas).

L'identité des amorces avec les 4 séquences de *MET1* listée plus haut a été vérifiée en utilisant l'outil BLAST (paramètres pas défaut). Les amorces présentant 100% d'identité avec les quatre séquences ont été conservées pour le crible des populations. Les 14 premiers marqueurs respectant ces critères ont été utilisés pour le crible (Annexes).

3.3. Crible des populations

Les populations de TILLING criblées sont :(i) une population d'un blé diploïde *T. monoccocum*, Pays de Sault (INRA) qui compte 1840 individus (l'équivalent de 20 plaques 96 puits avec 3 puits H₂O et 1 puit Pays de Sault par plaque) ; (ii) une population de blé hexaploïde *T. aestivum*, Arche (Biogemma) qui compte 912 individus et (iii) une population de blé hexaploïde *T. aestivum* Apache (Biogemma) qui compte 5350 individus.

Le crible est fait en utilisant la méthode KASP™ (LGC Genomics, <http://www.lgcgenomics.com/genotyping/>) (Fig 24).

Les conditions de mix PCR sont les suivantes :

mix d'amorces	concentration finale (μM)	mix PCR	volume pour 1 réaction (μl)
amorce A1 (100μM)	12	KASP V4.0 2X Master Mix	2
amorce A2 (100μM)	12	mix d'amorces	0,055
amorce C (100μM)	30	total	2,055
		vol mix/réaction	2
		volume ADN (2ng/μl)	2
		volume total réaction	4

Les amplifications sont réalisées dans un Hydrocycler (LGC Genomics) sur 41 cycles à 61°C. Si l'amplification n'est pas suffisante, des cycles supplémentaires sont effectués par série de 3 jusqu'à un maximum de 50 cycles totaux.

Les résultats de lecture de fluorescence sortent en fichier .sds et sont ensuite analysés avec l'outil Taqmangenotyper (Invitrogen).

3.4. Validation des mutants

Les individus qui ressortent lors de ce crible comme des mutants potentiels sont de nouveau testés en triplicat dans les mêmes conditions PCR, avec un individu sauvage comme témoin, en triplicat également. Si l'analyse de la fluorescence confirme qu'il s'agit de mutant, la dernière validation consiste en un séquençage.

4. Etablissement et validation des lignées RNAi

4.1. Clonage

Pour la première vague de production des lignées RNAi, la séquence du contig EST TA83026 a été utilisée pour définir un fragment de 500pb spécifique de *MET1*.

Pour la deuxième vague de production des lignées RNAi, la séquence de la copie du transcrit de *TaMET-7A1* a été utilisée pour définir un fragment de 630pb ; la localisation du fragment a été choisie pour cibler les neuf copies de *TaMET-1*.

Pour chaque vague de RNAi, les fragments définis ont été synthétisés en orientation sens et anti-sens séparés par une séquence intronique : StIS1.

Le clonage a été effectué pour les deux vagues de crible par la méthode du GoldenGate (Engler et al., 2009) qui s'effectue en une seule réaction.

Le clonage s'est fait avec le promoteur Ubiquitine du maïs (proZmUbi), avec le terminateur terAtNOS dans un vecteur de type pBIOS.

Le clonage pour la deuxième vague de crible a été réalisé par Biogemma en utilisant la même méthode (Annexes)

Les conditions de clonage sont les suivantes :

	volume (µl)
Promoteur (50ng/µl)	1
Gène (50ng/µl)	1
Termineur (50ng/µl)	1
Vecteur (50ng/µl)	1
Enzyme de restriction (5U/µl)	0,5
Ligase (3U/µl)	0,75
Tampon de ligase 10X	1
H2O	qsp 10

La réaction est placée 50min à 37°C.

5µl de cette réaction sont ajoutés à 25µl de NEB 5-alpha CompetentE.coli (New EnglandBiolabs). Le mélange est ensuite incubé 30min dans la glace, puis 40sec à 40°C et de nouveau 5min dans la glace. 300µl de milieu de SOC sont ajoutés et incubé 1h à 37°C sous agitation. Enfin la culture est étalée sur boîte LB+ampicilline+kanamycine et incubé à 37°C sur la nuit.

Les colonies qui ont poussé sont tout d'abord vérifiées par PCR. Plusieurs amplifications sont effectuées entre le promoteur et la séquence du RNAi, entre le RightBorder (RB) et le promoteur, et entre le terminateur et le vecteur.

Les colonies positives sont cultivées dans 70ml de LB + kanamycine+ ampicilline à 37°C sur la nuit. Les extractions de plasmides sont faites avec le kit midiprep Macherey-Nagel puis l'ADN dosé.

Une seconde vérification par digestion enzymatique est effectuée, pour la première vague avec deux couples d'enzymes : XhoI/XmnI et NcoI/SpeI.

Plusieurs stocks glycerol du clone est fait à partir de 10ml de culture auxquels sont ajoutés 2ml de glycerol.

Enfin l'ADN du clone positif est envoyé au séquençage pour dernière vérification.

4.2. Transformation Agrobacterium

A partir du vecteur sélectionné plus haut la transformation en *Agrobacterium* est effectuée comme suit : 25ng d'ADN contenu dans 30µl d'H₂O sont ajoutés à 2µl de bactéries. La transformation est faite par electroporation puis 300µl de milieu SOC sont ajoutés et incubés 5h à 28°C. La culture est ensuite étalée sur boîte LB + rifampicine + kanamycine.

Deux colonies sont cultivées dans du milieu YT liquide+ rifampicine + kanamycine pendant 2 jours.

Validation de ces clones par amplification PCR spécifiques à *Agrobacterium* ou spécifique à la séquence insérée. De nouveaux stocks glycerol sont fait pour la souche qui sera utilisé pour les transformations des plantes.

4.3. Transformation du blé et génotypage des évènements

La transformation du blé est faite selon la méthode de la Seed Inoculation Method (SIM) décrite dans (Ishida *et al.*, 1996).

Après régénération des plantes, l'intégrité du transgène est vérifiée par une PCR au niveau du promoteur ZmUbi. Le nombre de copies insérées et la recherche d'homozygotes et effectué par RT-qPCR sur le gène marqueurs NptII.

4.4. Suivi des générations des plantes issues de la 1ère vague RNAi

7 évènements indépendants d'insertion ont été détectés à la génération T0. Certaines insertions se trouvent dans plusieurs plantes sœurs et au total 20 plantes T0 sont issues de ces 7 évènements (Tableau 10). Sur ces 20 plantes, 6 n'ont donné aucune graine T1, 6 autres, moins de 10 graines et parmi les 8 restantes seule une a donné plus de 100 graines; pour les 7 autres, on dénombre entre 14 et 28 graines. Après génotypage par Biogemma, seules 3 plantes T0 sur 20 ont un transgène intègre c'est-à-dire sans réorganisation et une seule a donné une descendance T1.

Du fait du faible nombre de plantes récupérées en T0, toutes les graines obtenues ont été conservées quelque soit les résultats d'intégrité du construit. Pour 8 des plantes T0, dont la lignée intègre et deux lignées ayant un transgène tronqué, 14 à 16 graines T1 ont été semées : B0052259, B0052439, B0052436, B0052441, B0052442, B0053037, B0053038, B0053041.

Aucune des graines issues de la lignée intègre, B0053041, n'ont germé. Pour toutes les autres lignées T0, entre 2 et 14 plantes T1 ont germé et toutes ont donné des graines T2.

Seules des graines issues de la lignée T0 ont été B0053038 ont été semées pour obtenir des plantes de génération T2 (Tableau 6).

4.5. Suivi des générations des plantes issues de la 2ème vague RNAi

22 évènements T0 indépendants ont été obtenus mais comme lors de la première vague, une forte stérilité a été observée et seules 12 plantes ont donné des descendants (Tableau 11). Sur ces 12 plantes T0, 8 sont positives pour l'insertion d'un transgène intègre. Pour ces 12 évènements, entre 6 et 16 graines T1 ont été semées. Après génotypage des plantes T1, la descendance de 3 lignées T0 a été éliminée car aucun homozygote ou hémizygote n'a été retrouvé. Toutes les autres lignées ont été conservées pour obtenir des graines T2.

nom plante T0	n° plante / cal	nombre de graines	intégrité du construit	nombre de graines T1 ressemées	nombre de graines T1 germées
B0052259	CW-00992A-1	131		14	8
B0052433	CW-00992A-10	3			
B0052434	CW-00992A-11	1			
B0052435	CW-00992A-12	7			
B0052439	CW-00992A-2	21		16	3
B0052436	CW-00992A-3	24		16	2
B0052441	CW-00992A-4	28		16	7
B0052442	CW-00992A-5	20		16	6
B0052587	CW-00992A-6				
B0052437	CW-00992A-7	5			
B0052440	CW-00992A-8	1			
B0052438	CW-00992A-9	2			
B0052260	CW-00992B-1		positif		
B0053479	CW-00992B-2				
B0053480	CW-00992B-4				
B0053037	CW-00994A-1	25	tronqué	15	14
B0053038	CW-00994B-1	24	tronqué	16	12
B0053039	CW-00994C-1				
B0053040	CW-00994D-1		positif		
B0053041	CW-00994E-1	14	positif	14	0

Tableau 10 : Effectifs des plantes issues de la première vague de RNAi

Le suivi est donné jusqu'à la génération T1.

nom plante T0	n° plante / cal	nombre de graines T1	intégrité du construit	nombre de graines T1 ressemées	nombre de graines T1 germées	
B0075111	CW-01460B-1	99	positif	15	14	
B0075224	CW-01461B-1	81		16	15	Éliminée
B0075228	CW-01461F-1	6	positif	6	5	
B0075232	CW-01461J-1	35	positif	16	9	
B0075483	CW-01460C-1	17	positif	16	6	
B0075961	CW-01459B-1	59		16	5	Éliminée
B0075962	CW-01459C-1	74	positif	16	8	
B0075963	CW-01459D-1	87		16	10	
B0075965	CW-01459F-1	74		14	11	Éliminée
B0075966	CW-01460D-1	73	positif	16	15	
B0075968	CW-01460F-1	69	positif	16	10	
B0075969	CW-01460G-1	19	positif	15	12	

Les traits horizontaux délimitent les plantes sœurs issues d'un même évènement de transformation.

Tableau 11 : Effectifs des plantes issues de la deuxième vague de RNAi

Le suivi est donné jusqu'à la génération T1. Après génotypage des plantes T1, celles issues de trois des lignées T0 (B0075224, B0075961 et B0075965) étaient toutes négatives pour la présence d'un transgène. Elles ont été éliminées.

4.6. Recherche des sRNA par Northern Blot

Les ARN totaux ont été extraits de plantes T1 de la première vague de RNAi âgées de deux mois, selon la méthode adaptée de (Bogorad *et al.*, 1983). Ces ARNs ont été dosés au Nanodrop et leur concentration ajustée à 1000 ng/ μ l ou 1400 ng/ μ l selon les prélèvements. Des aliquots de 15 μ g ont ensuite été réalisés et conservés à -80°C.

Le gel de migration à 15% d'acrylamide a été préparé avec le kit Rotiphorese Système de séquençage de l'ADN (Roth-Sochiel) selon les instructions du fournisseur.

A un volume V d'ARN est ajouté 1,5V de tampon de charge préparé comme suit: 9,8ml de formamide + 2,5 à 10mg d'un mélange xylène cyanole et bromophénol + 200 μ l d'EDTA 0,5M pH8. Après une incubation de 5min à 95°C puis de 3min dans la glace, les échantillons sont déposés dans le gel et la migration s'effectue comme suit : 30 à 45 min à 3W puis 18W jusqu'à la sortie de la bande basse de bleu de bromophénol du gel. Le transfert se fait ensuite sur membrane Hybond-NX sur la nuit à 0,2A.

La fixation chimique se fait dans une solution d'EDC préparée comme suit: 12,5M de 1-méthylimidazole dans 10ml d'eau RNA-free, après ajustement à pH8 avec de l'HCl 1N, ajout de 0,373g d'EDC et enfin ajustement du volume à 12ml. La membrane est incubée dans cette solution 1h30 à 60°C puis rincée trois fois à l'eau et séchée à l'air libre. La conservation se fait dans du film plastique sous vide à -20°C.

Pour la révélation radioactive, les sondes sont préparées comme suit : 10pmol oligonucleotides + 2 μ l tampon kinase + 5 μ l γ ATP (P32, marqué à 6000 Curie/mol) + 1,5 μ l kinase + H₂O qsp 20 μ l. Incubation d'1h à 37°C, purification sur colonne et élution dans 100 μ l.

Les membranes sont tout d'abord préhybridées dans une solution préparée comme suit : tampon 5X SSC + 20mM NaH₂PO₄ pH7.2 + 7% SDS + 2X Denhardt's solution + 40 à 50 μ g/ml ADN de sperme de saumon ; incubation 4h dans une étuve à 50°C.

La solution de sondes est ensuite ajoutée à la membrane, présente dans le tampon de préhybridation. Après une incubation sur la nuit à 50°C, la membrane est rincée deux fois 10 min dans un mélange : 3X SSC + 25mM NaH₂PO₄ pH7.5 + 5% SDS ; puis 1 fois 10 min dans un mélange : 1X SSC + 1% SDS.

La membrane est autoradiographiée sur un film pour une durée de 1 journée pour les sondes ciblant les siRNA témoins, et 3 jours pour les sondes ciblant les séquences des siRNA MET1.

Nom	Séquence	taille de l'oligonucléotide
oligo-16nt_sde-MET1-1	CTTAGATATGTGATCG	16
oligo-21nt_sde-MET1-2	CTTCCTCCCAAGCGCAAAGGC	21
oligo-25nt_sde-MET1-1	TAATTCATTCATCTCCTTAGATATG	25
oligo-30nt_sde-MET1-2	CGACGGCTTCCTTCAGCTTCCTCCCAAGCG	30

Tableau 12 : Oligonucléotides utilisés comme échelle externe pour les Northern Blot

Nom	Séquence	taille de la sonde	taille du sRNA
sonde 1 MET1	CGATCACATATCTAAGGAGATGAATGAATTAATC	35	
sonde 2 MET1	TTGCCTTTGCGCTTGGGAGGAAGCTGAAGGAAGCCGTCGATG	42	
miRNA 159	AAGAGCTCCCTTCAATCCAAA	21	21
sonde U6	AGGGGCCATGCTAATCTTCTC	21	110

Tableau 13 : Sondes utilisées pour l'hybridation des membranes

Les sondes 1 et 2 permettent de révéler les siRNA ciblant le gène *MET1*. Les deux autres sondes permettent de révéler l'ARN U6 et le miRNA159.

Un mélange d'oligonucléotides de taille différentes (16, 21, 25, 30 nt) a été utilisé comme échelle externe (Tableau 12) ; sur chaque gel une quantité de 10 pg et 1 pg a été déposée, ainsi qu'un témoin négatif NB1 (lignée sauvage).

Deux sondes ont été utilisées, en mélange, pour reconnaître les siRNA produit par le RNAi. Ces deux sondes vont également reconnaître les oligonucléotides de l'échelle externe (Tableau 13).

Le siRNA U6 et le miRNA 159 ont été migrés sur gel comme témoin interne d'hybridation et de taille. Les sondes utilisées pour l'hybridation de ces deux ARN ont été utilisées en mélange.

5. Etude des profils de méthylation

5.1. Extraction de l'ADNg

Le matériel végétal est broyé manuellement au mortier pilon dans de l'azote liquide. L'ADNg est extrait avec les kits Qiagen : DNeasy Plant Mini/Maxi kit en suivant les instructions données par le fournisseur. La quantité de matériel de départ est de 100mg pour le Mini kit et de 1g pour le Maxi.

Dans le cas d'une extraction avec le Maxi kit, les deux éluations sont réalisées dans un volume de 500µl chacune et ne sont pas mélangées : éluat 1 et éluat 2.

Dans le cas d'une extraction avec le Mini kit, les deux éluations sont réalisées dans un volume de 100µl chacune et sont mélangées. Les 200µl d'ADNg sont ensuite précipités à l'éthanol et repris dans 100µl.

Les ADNg sont dosés au nanodrop, avec le tampon d'éluat qui sert de blanc.

En parallèle, 10µl d'une dilution au 1/10^e des ADNg additionné de 2µl de bleu bromophénol 6X, éluat 1 et éluat 2, sont déposés sur un gel 1% avec une gamme étalon de phage λ qui compte 4-5 points (15, 31, 63, 125, 250 ng).

5.2. Crible enzymatique des méthylations

Crible enzymatique des gènes cibles avec McrBC, HpaII et MspI, réalisé sur des ADNg sur l'éluat 2 d'extraction d'ADNg sur Maxi kit.

		McrBC		HpaII		MspI	
		+	-	+	-	+	-
Matrice		250ng	250ng	250ng	250ng	250ng	250ng
Tampon	10X	2µl	2µl	2µl	2µl	2µl	2µl
Enzyme (NEB)	10 000 U/ml	0,5µl		0,5µl		0,5µl	
BSA	20X	1µl	1µl				
1mM GTP	20X	1µl	1µl				
H2O		qsp 20µl	qsp 20µl	qsp 20µl	qsp 20µl	qsp 20µl	qsp 20µl

Tableau 14 : Conditions de digestion des enzymes sensibles à la méthylation

Les digestions sont laissées sur la nuit à 37°C selon les conditions décrite dans le tableau 14.

Après la digestion, de l'eau est ajoutée pour compléter le volume à 100µl. L'échantillon obtenu correspond à la matrice utilisé pour les amplifications PCRs après digestion selon les conditions suivantes :

Mix PCR	Concentration initiale	Volume (µl)
Green GoTaq® Reaction Buffer	5X	4
MgCl ₂	25mM	1,5
dNTPs	2,5mM chaque	1
Amorce sens	10µM	1
Amorce anti sens	10µM	1
Matrice		5µl
H ₂ O		qsp 20µl

Programme d'amplification :

95°C	2min	
95°C	30 sec	↘1°C sur 10 cycles
60°C à 53°C	30 sec	
72°C	45sec	
95°C	30 sec	15 cycles
53°C	30 sec	
72°C	45 sec	
72°C	5 min	
15°C	pause	

Les couples utilisés pour les amplifications PCR après digestion enzymatique sont donnés en annexes.

Les produits PCRs sont déposés sur un gel agarose 1%, BET.

5.3. Conversion au bisulfite

La conversion au bisulfite est réalisée avec le kit EZ DNA Methylation-Gold™ kit (ZymoResearch) en suivant les instructions données par le fournisseur. Les extractions sur les Mini kit ont été utilisées comme matrices.

5.4. Design des amorces pour l'amplification sur ADN converti au bisulfite pour un séquençage MiSeq (Illumina)

5.4.1. Séquences utilisées pour le design

Biogemma nous a dans un premier temps donné accès aux séquences des gènes *Vrn-A1*, *Vrn-A2*, *Vrn-D2*, *Rht-A1-prom*, *Rht-A1*, *Rht-B1*, *Rht-D1*, *Ppd-A1*, *Ppd-B1*, *Ppd-D1*, *Vrn-B3*, *Fta*, *Ftd*. Pour être sûr de disposer de toute la séquence du gène et afin d'obtenir les séquences homologues manquantes pour certains d'entre eux, une recherche par BLASTn a été réalisée entre ces séquences candidates et les données génomiques IWGSC.

Des séquences supplémentaires ont été identifiées pour *Vrn1* et *Vrn3-FT*, elles ont été utilisées en complément pour la suite des analyses.

5.4.2. Etablissement des profils CpG

A partir des séquences des gènes cible, alignement par le logiciel MUSCLE (paramètres par défaut) (Edgar, 2004). Utilisation du script perl pour établir les profils CpG, chapitre 1.

5.4.3. Règles pour la définition des amorces

Les amorces utilisées pour l'amplification sur ADN converti au bisulfite (ADN-BS) sont de types dégénérées : c'est-à-dire que l'on retrouve des bases dégénérées dans leur séquence, Y ou R selon que l'on soit dans l'amorces sens ou anti-sens.

Les amorces ont été dessinées manuellement selon les règles suivantes :

- Eviter les CpG et les CpHpG le plus possible
- Un maximum de trois cytosines par amorces, et si pour l'une ce n'est pas possible, pas plus de 5 cytosines dans les deux réunies
- Pas de cytosine dans les deux dernières bases en 3' des amorces
- Prendre des zones plutôt riches en guanine
- T_m entre 60 et 62°C, Oligo Analyzer 1.2 – pas de différence de plus de 2°C entre l'amorce sens et anti-sens.

Pour l'amorce sens, les cytosines, qui peuvent être des cytosines ou des thymines après conversion, sont remplacées par des Y.

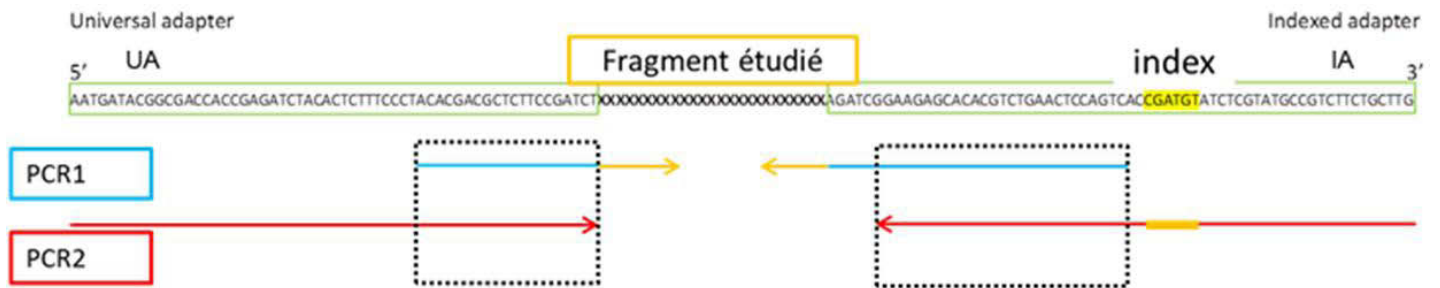


Figure 41 : Principe d'amplification par PCR imbriquée pour le séquençage MiSeq

L'adaptateur universel (UA) fait 58 bases et l'adaptateur indexé (IA) 64 bases (encadrés en vert).

La première PCR (PCR1) est faite avec des amorces incluant une zone spécifique de la cible de l'ordre de 20 bases (indiquée en jaune) à laquelle une première partie des adaptateurs est introduite (indiquée en bleu) : 20 bases pour l'UA et 32 bases pour l'IA. Pour ce dernier, la partie ajoutée ne contient pas l'index.

La 2^{ème} PCR (PCR2) utilise des amorces (indiquées en rouge) qui permettent de compléter les adaptateurs et d'inclure l'index (souligné en jaune dans la séquence, orange sur l'amorce). En fin de PCR, l'amorce UA est la même pour toutes les PCR2 alors l'amorce IA inclue un index spécifique du génotype étudié.

Pour l'amorce anti-sens, les guanines, qui seront en face de cytosine ou de thymine dans le fragment à amplifier, sont remplacées par des R.

La localisation des amorces a été choisie à partir des résultats du crible enzymatique et des profils de CpG, en ciblant plutôt des zones communes aux trois génomes.

Une fois les amorces spécifiques à chaque cible définies, la première partie des adaptateurs pour le séquençage MiSeq a été ajoutée aux amorces (Fig 41). La séquence (ACACGACGCTCTTCCGATCT) est ajoutée en 5' de l'amorce sens et la séquence (GACTGGAGTTCAGACGTGTGCTCTTCCGATCT) en 5' de l'amorce anti-sens.

En PCR2, l'amorce commune à toutes les amplifications, universal adapter (UA) est (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT) et l'amorce indexée, indexed adapter (IA), est (CAAGCAGAAGACGGCATACGAGATNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCC) avec N symbolisant la localisation de l'index de 6 nucleotides.

5.5. Amplification PCR sur ADN converti au bisulfite

Les amplifications après conversion au bisulfite pour la production des amplicons pour le séquençage MiSeq ont été réalisées dans les conditions suivantes :

Mix PCR :	Concentration initiale	Volume (µl)
AmpliTaq Gold® 360 PCR Master Mix	2X	10
amorce sens	10µM	1
amorce anti-sens	10µM	1
ADN-BS		1,5
H2O		qsp 20µl

Programme d'amplification :

95°C	10 min	
94°C	1 min	
65°C à 55°C	1 min	
72°C	1 min 30	∇1°C sur 10 cycles
94°C	1 min	
55°C à 50°C	1 min	
72°C	1 min 30	∇0,5°C sur 10 cycles
94°C	40 sec	
50°C	40 sec	
72°C	1 min	20 cycles
72°C	7 min	
15°C	pause	

Références

- Amado, L., Abranches, R., Neves, N., and Viegas, W. (1997). Development-dependent inheritance of 5-azacytidine-induced epimutations in triticale: analysis of rDNA expression patterns. *Chromosome Res.* *5*, 445–450.
- Anderson, J.A. (2007). Marker-assisted selection for Fusarium head blight resistance in wheat. *Int. J. Food Microbiol.* *119*, 51–53.
- Aufsatz, W., Mette, M.F., Matzke, A.J.M., and Matzke, M. (2004). The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides. *Plant Mol. Biol.* *54*, 793–804.
- Bartee, L., and Bender, J. (2001). Two Arabidopsis methylation-deficiency mutations confer only partial effects on a methylated endogenous gene family. *Nucleic Acids Res.* *29*, 2127.
- Baulcombe, D. (2004). RNA silencing in plants. *Nature* *431*, 356–363.
- Bennett, S. (2004). Solexa Ltd. *Pharmacogenomics* *5*, 433–438.
- 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.
- Berkman, P.J., Visendi, P., Lee, H.C., Stiller, J., Manoli, S., Lorenc, M.T., Lai, K., Batley, J., Fleury, D., Šimková, H., et al. (2013). Dispersion and domestication shaped the genome of bread wheat. *Plant Biotechnol. J.* *11*, 564–571.
- Bernacchia, G., Primo, A., Giorgetti, L., Pitto, L., and Cella, R. (1998). Carrot DNA-methyltransferase is encoded by two classes of genes with differing patterns of expression. *Plant J.* *13*, 317–329.
- Bervini, S., and Herzog, H. (2013). Mouse models of Prader–Willi Syndrome: A systematic review. *Front. Neuroendocrinol.* *34*, 107–119.
- Bian, Q., and Belmont, A.S. (2012). Revisiting higher-order and large-scale chromatin organization. *Curr. Opin. Cell Biol.* *24*, 359–366.
- Bogorad, L., Gubbins, E.J., Krebbers, E., Larrinua, I.M., Mulligan, B.J., Muskavitch, K.M.T., Orr, E.A., Rodermel, S.R., Schantz, R., Steinmetz, A.A., et al. (1983). Cloning and physical mapping of maize plastid genes. In *Methods in Enzymology*, B.F. Sidney Fleischer, ed. (Academic Press), pp. 524–554.
- Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science* *317*, 1760–1764.
- Boutet, S., Vazquez, F., Liu, J., Béclin, C., Fagard, M., Gratias, A., Morel, J.-B., Crété, P., Chen, X., and Vaucheret, H. (2003). Arabidopsis HEN1: A Genetic Link between Endogenous miRNA Controlling Development and siRNA Controlling Transgene Silencing and Virus Resistance. *Curr. Biol.* *13*, 843–848.

- Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L.A., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., et al. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* *491*, 705–710.
- Brisson, N., Gate, P., Gouache, D., Charmet, G., Oury, F.-X., and Huard, F. (2010). Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *Field Crops Res.* *119*, 201–212.
- Buard, J., and de Massy, B. (2007). Playing hide and seek with mammalian meiotic crossover hotspots. *Trends Genet.* *23*, 301–309.
- Callebaut, I., Courvalin, J.-C., and Mornon, J.-P. (1999). The BAH (bromo-adjacent homology) domain: a link between DNA methylation, replication and transcriptional regulation. *FEBS Lett.* *446*, 189–193.
- Charmet, G. (2011). Wheat domestication: Lessons for the future. *C. R. Biol.* *334*, 212–220.
- Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S., and Peacock, W.J. (1997). Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 4223–4228.
- Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003). Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* *23*, 5594–5605.
- Chen, X., Chen, Z., Zhao, H., Zhao, Y., Cheng, B., and Xiang, Y. (2014). Genome-Wide Analysis of Soybean HD-Zip Gene Family and Expression Profiling under Salinity and Drought Treatments. *PloS One* *9*, e87156.
- Cheng, Z., Buell, C.R., Wing, R.A., Gu, M., and Jiang, J. (2001). Toward a Cytological Characterization of the Rice Genome. *Genome Res.* *11*, 2133–2141.
- Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E., and Fischer, R.L. (2002). DEMETER, a DNA Glycosylase Domain Protein, Is Required for Endosperm Gene Imprinting and Seed Viability in *Arabidopsis*. *Cell* *110*, 33–42.
- Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., Schlub, S., Le Paslier, M.-C., Magdelenat, G., Gonthier, C., et al. (2010). Megabase Level Sequencing Reveals Contrasted Organization and Evolution Patterns of the Wheat Gene and Transposable Element Spaces[W]. *Plant Cell* *22*, 1686–1701.
- 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 bisulfite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* *452*, 215–219.
- Colomé-Tatché, M., Cortijo, S., Wardenaar, R., Morgado, L., Lahouze, B., Sarazin, A., Etcheverry, M., Martin, A., Feng, S., Duvernois-Berthet, E., et al. (2012). Features of the *Arabidopsis* recombination landscape resulting from the combined loss of sequence variation and DNA methylation. *Proc. Natl. Acad. Sci.* *109*, 16240–16245.

- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. *Cell* 146, 67–79.
- Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* 25, 1010–1022.
- Van Driel, R., and Fransz, P. (2004). Nuclear architecture and genome functioning in plants and animals: what can we learn from both? *Exp. Cell Res.* 296, 86–90.
- Eamens, A., Wang, M.-B., Smith, N.A., and Waterhouse, P.M. (2008). RNA Silencing in Plants: Yesterday, Today, and Tomorrow. *Plant Physiol.* 147, 456–468.
- Ebbs, M.L., and Bender, J. (2006). Locus-specific control of DNA methylation by the Arabidopsis SUVH5 histone methyltransferase. *Plant Cell* 18, 1166–1176.
- 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, 10507–10515.
- Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res.* 10, 2709–2721.
- Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type II Restriction Enzymes. *PLoS ONE* 4, e5553.
- Esteller, M. (2008). Epigenetics in Cancer. *N. Engl. J. Med.* 358, 1148–1159.
- Fagard, M., and Vaucheret, H. (2000). (TRANS)GENE SILENCING IN PLANTS: How Many Mechanisms? *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 51, 167–194.
- Feng, S., and Jacobsen, S.E. (2011). Epigenetic modifications in plants: an evolutionary perspective. *Curr. Opin. Plant Biol.* 14, 179–186.
- Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci.* 107, 8689–8694.
- Finnegan, E.J., Peacock, W.J., and Dennis, E.S. (1996). Reduced DNA methylation in Arabidopsis thaliana results in abnormal plant development. *Proc. Natl. Acad. Sci.* 93, 8449–8454.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811.

- Florea, L., Hartzell, G., Zhang, Z., Rubin, G.M., and Miller, W. (1998). A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* *8*, 967–974.
- Fransz, P., Jong, J.H. de, 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.* *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.
- Freeman, J., Sparks, C.A., West, J., Shewry, P.R., and Jones, H.D. (2011). Temporal and spatial control of transgene expression using a heat-inducible promoter in transgenic wheat. *Plant Biotechnol. J.* *9*, 788–796.
- Fu, D., Uauy, C., Blechl, A., and Dubcovsky, J. (2007). RNA interference for wheat functional gene analysis. *Transgenic Res.* *16*, 689–701.
- Fujimoto, R., Sasaki, T., and Nishio, T. (2006). Characterization of DNA methyltransferase genes in *Brassica rapa*. *Genes Genet. Syst.* *81*, 235–242.
- Fussner, E., Ching, R.W., and Bazett-Jones, D.P. (2011). Living without 30 nm chromatin fibers. *Trends Biochem. Sci.* *36*, 1–6.
- Gao, F., Liu, X., Wu, X.-P., Wang, X.-L., Gong, D., Lu, H., Xia, Y., Song, Y., Wang, J., Du, J., et al. (2012). Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*. *Genome Biol.* *13*, R100.
- Gehring, M., Huh, J.H., Hsieh, T.-F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (2006). DEMETER DNA Glycosylase Establishes MEDEA Polycomb Gene Self-Imprinting by Allele-Specific Demethylation. *Cell* *124*, 495–506.
- Gehring, M., Missirian, V., and Henikoff, S. (2011). Genomic Analysis of Parent-of-Origin Allelic Expression in *Arabidopsis thaliana* Seeds. *PLoS ONE* *6*, e23687.
- Geyer, P.K., Vitalini, M.W., and Wallrath, L.L. (2011). Nuclear organization: taking a position on gene expression. *Curr. Opin. Cell Biol.* *23*, 354–359.
- Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldán-Arjona, T., David, L., and Zhu, J.-K. (2002). ROS1, a Repressor of Transcriptional Gene Silencing in *Arabidopsis*, Encodes a DNA Glycosylase/Lyase. *Cell* *111*, 803–814.
- Grossniklaus, U., Vielle-Calzada, J.P., Hoepfner, M.A., and Gagliano, W.B. (1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science* *280*, 446–450.
- Grossniklaus, U., Spillane, C., Page, D.R., and Köhler, C. (2001). Genomic imprinting and seed development: endosperm formation with and without sex. *Curr. Opin. Plant Biol.* *4*, 21–27.

- Gruenbaum, Y., Naveh-Many, T., Cedar, H., and Razin, A. (1981). Sequence specificity of methylation in higher plant DNA. *Nature* 292, 860–862.
- Guo, J.U., Su, Y., Zhong, C., Ming, G., and Song, H. (2011). Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell* 145, 423–434.
- Hedden, P. (2003). The genes of the Green Revolution. *Trends Genet.* 19, 5–9.
- Heitz, E. (1928). Das Heterochromatin der Moose. *Jahrb. Für Wiss. Bot.* 69, 762–818.
- Herman, J.G., Merlo, A., Mao, L., Lapidus, R.G., Issa, J.-P.J., Davidson, N.E., Sidransky, D., and Baylin, S.B. (1995). Inactivation of the CDKN2/p16/MTS1 Gene Is Frequently Associated with Aberrant DNA Methylation in All Common Human Cancers. *Cancer Res.* 55, 4525–4530.
- Herman, J.G., Graff, J.R., Myöhänen, S., Nelkin, B.D., and Baylin, S.B. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. U. S. A.* 93, 9821–9826.
- Howard, G., Eiges, R., Gaudet, F., Jaenisch, R., and Eden, A. (2007). Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* 27, 404–408.
- Hsieh, T.-F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., et al. (2011). Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc. Natl. Acad. Sci.* 108, 1755–1762.
- Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG Islands Identify Numerous Conserved Promoters in the Mammalian Genome. *PLoS Genet.* 6, e1001134.
- International Rice Genome Sequencing (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800.
- Ishida, Y., Saito, H., Ohta, S., Hiei, Y., Komari, T., and Kumashiro, T. (1996). High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*. *Nat. Biotechnol.* 14, 745–750.
- Jacobsen, S.E., Sakai, H., Finnegan, E.J., Cao, X., and Meyerowitz, E.M. (2000). Ectopic hypermethylation of flower-specific genes in Arabidopsis. *Curr. Biol.* 10, 179–186.
- Jasencakova, Z., Meister, A., and Schubert, I. (2001). Chromatin organization and its relation to replication and histone acetylation during the cell cycle in barley. *Chromosoma* 110, 83–92.
- Jeddeloh, J.A., Stokes, T.L., and Richards, E.J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat. Genet.* 22, 94–97.
- Jeltsch, A. (2010). Phylogeny of Methylomes. *Science* 328, 837–838.

- Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuissou, J., Heredia, F., Audigier, P., et al. (2009a). Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits. *PLoS Genet.* 5, e1000530.
- Johannes, F., Porcher, E., Teixeira, F.K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuissou, J., Heredia, F., Audigier, P., et al. (2009b). Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet.* 5, e1000530.
- Kakutani, T., Jeddelloh, J.A., and Richards, E.J. (1995). Characterization of an *Arabidopsis thaliana* DNA hypomethylation mutant. *Nucleic Acids Res.* 23, 130.
- Kankel, M.W., Ramsey, D.E., Stokes, T.L., Flowers, S.K., Haag, J.R., Jeddelloh, J.A., Riddle, N.C., Verbsky, M.L., and Richards, E.J. (2003). *Arabidopsis* MET1 Cytosine Methyltransferase Mutants. *Genetics* 163, 1109–1122.
- Kapazoglou, A., Drosou, V., Argiriou, A., and Tsafaris, A.S. (2013). The study of a barley epigenetic regulator, HvDME, in seed development and under drought. *BMC Plant Biol.* 13, 172.
- Keller, T.E., and Yi, S.V. (2014a). DNA methylation and evolution of duplicate genes. *Proc. Natl. Acad. Sci.* 201321420.
- Keller, T.E., and Yi, S.V. (2014b). DNA methylation and evolution of duplicate genes. *Proc. Natl. Acad. Sci. U. S. A.*
- Kellis, M., Birren, B.W., and Lander, E.S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428, 617–624.
- Khan, A.R., Enjalbert, J., Marsollier, A.-C., Rousselet, A., Goldringer, I., and Vitte, C. (2013). Vernalization treatment induces site-specific DNA hypermethylation at the VERNALIZATION-A1 (VRN-A1) locus in hexaploid winter wheat. *BMC Plant Biol.* 13, 209.
- Kim, M., Ohr, H., Lee, J.W., Hyun, Y., Fischer, R.L., and Choi, Y. (2008). Temporal and spatial downregulation of *Arabidopsis* MET1 activity results in global DNA hypomethylation and developmental defects. *Mol. Cells* 26, 611–615.
- Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-Way Control of FWA Imprinting in *Arabidopsis* Endosperm by DNA Methylation. *Science* 303, 521–523.
- Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W.J.J., Koornneef, M., and Kakutani, T. (2007). Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J.* 49, 38–45.
- Kovalchuk, N., Jia, W., Eini, O., Morran, S., Pyvovarenko, T., Fletcher, S., Bazanova, N., Harris, J., Beck-Oldach, K., Shavrukov, Y., et al. (2013). Optimization of TaDREB3 gene expression in transgenic barley using cold-inducible promoters. *Plant Biotechnol. J.* 11, 659–670.

- Lai, J., Ma, J., Swigoňová, Z., Ramakrishna, W., Linton, E., Llaca, V., Tanyolac, B., Park, Y.-J., Jeong, O.-Y., Bennetzen, J.L., et al. (2004). Gene Loss and Movement in the Maize Genome. *Genome Res.* *14*, 1924–1931.
- Laird, P.W. (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nat. Rev. Genet.*
- 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.
- Lawton-Rauh, A. (2003). Evolutionary dynamics of duplicated genes in plants. *Mol. Phylogenet. Evol.* *29*, 396–409.
- Lindroth, A.M., Cao, X., Jackson, J.P., Zilberman, D., McCallum, C.M., Henikoff, S., and Jacobsen, S.E. (2001). Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation. *Science* *292*, 2077–2080.
- Lindroth, A.M., Shultis, D., Jasencakova, Z., Fuchs, J., Johnson, L., Schubert, D., Patnaik, D., Pradhan, S., Goodrich, J., Schubert, I., et al. (2004). Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J.* *23*, 4286–4296.
- Ling, H.-Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C., Fan, H., Li, D., Dong, L., Tao, Y., et al. (2013). Draft genome of the wheat A-genome progenitor *Triticum urartu*. *Nature* *496*, 87–90.
- Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* *430*, 471–476.
- Lister, R., and Ecker, J.R. (2009). Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res.* *19*, 959–966.
- Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly Integrated Single-Base Resolution Maps of the Epigenome in *Arabidopsis*. *Cell* *133*, 523–536.
- Lister, R., Pelizzola, M., Downen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* *462*, 315–322.
- Liu, X., Yu, C.-W., Duan, J., Luo, M., Wang, K., Tian, G., Cui, Y., and Wu, K. (2012). HDA6 Directly Interacts with DNA Methyltransferase MET1 and Maintains Transposable Element Silencing in *Arabidopsis*. *Plant Physiol.* *158*, 119–129.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* *389*, 251–260.
- Luo, M., Taylor, J.M., Spriggs, A., Zhang, H., Wu, X., Russell, S., Singh, M., and Koltunow, A. (2011). A Genome-Wide Survey of Imprinted Genes in Rice Seeds Reveals Imprinting Primarily Occurs in the Endosperm. *PLoS Genet* *7*, e1002125.

- Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* *408*, 538–540.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bembien, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005a). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* *437*, 376–380.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bembien, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005b). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* *437*, 376–380.
- 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.
- McCallum, C.M., Comai, L., Greene, E.A., and Henikoff, S. (2000). Targeted screening for induced mutations. *Nat. Biotechnol.* *18*, 455–457.
- Meyer, P., Niedenhof, I., and Lohuis, M. ten (1994). Evidence for cytosine methylation of non-symmetrical sequences in transgenic *Petunia hybrida*. *EMBO J.* *13*, 2084.
- Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J., and Mathieu, O. (2009a). Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* *461*, 427–430.
- Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J., and Mathieu, O. (2009b). Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* *461*, 427–430.
- Mirouze, M., Lieberman-Lazarovich, M., Aversano, R., Bucher, E., Nicolet, J., Reinders, J., and Paszkowski, J. (2012). Loss of DNA methylation affects the recombination landscape in Arabidopsis. *Proc. Natl. Acad. Sci.* *109*, 5880–5885.
- Molnar, A., Melnyk, C., and Baulcombe, D.C. (2011). Silencing signals in plants: a long journey for small RNAs. *Genome Biol.* *12*, 215.
- Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmeler, K., Hause, G., Aurich, A.-C., Dorn, R., Jenuwein, T., et al. (2005). Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis. *EMBO J.* *24*, 1418–1429.
- 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*, 977–979.
- Ohad, N., Margossian, L., Hsu, Y.C., Williams, C., Repetti, P., and Fischer, R.L. (1996). A mutation that allows endosperm development without fertilization. *Proc. Natl. Acad. Sci. U. S. A.* *93*, 5319–5324.
- Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* *99*, 247–257.

- Ortega-Galisteo, A.P., Morales-Ruiz, T., Ariza, R.R., and Roldán-Arjona, T. (2008). Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Mol. Biol.* *67*, 671–681.
- Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., et al. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* *457*, 551–556.
- Pearce, S., Saville, R., Vaughan, S.P., Chandler, P.M., Wilhelm, E.P., Sparks, C.A., Al-Kaff, N., Korolev, A., Boulton, M.I., Phillips, A.L., et al. (2011). Molecular Characterization of Rht-1 Dwarfing Genes in Hexaploid Wheat. *PLANT Physiol.* *157*, 1820–1831.
- Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007). DNA demethylation in the Arabidopsis genome. *Proc. Natl. Acad. Sci.* *104*, 6752–6757.
- Pradhan, M., Estève, P.-O., Chin, H.G., Samaranayake, M., Kim, G.-D., and Pradhan, S. (2008). CXXC domain of human DNMT1 is essential for enzymatic activity. *Biochemistry (Mosc.)* *47*, 10000–10009.
- Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. *Nat. Rev. Mol. Cell Biol.* *10*, 192–206.
- Rangwala, S.H., and Richards, E.J. (2007). Differential Epigenetic Regulation Within an Arabidopsis Retroposon Family. *Genetics* *176*, 151–160.
- Ratel, D., Ravanat, J.-L., Berger, F., and Wion, D. (2006). N6-methyladenine: the other methylated base of DNA. *BioEssays* *28*, 309–315.
- Ream, T.S., Woods, D.P., and Amasino, R.M. (2012). The Molecular Basis of Vernalization in Different Plant Groups. *Cold Spring Harb. Symp. Quant. Biol.* *77*, 105–115.
- Reinders, J., Wulff, B.B.H., Mirouze, M., Marí-Ordóñez, A., Dapp, M., Rozhon, W., Bucher, E., Theiler, G., and Paszkowski, J. (2009a). Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. *Genes Dev.* *23*, 939–950.
- Reinders, J., Wulff, B.B.H., Mirouze, M., Marí-Ordóñez, A., Dapp, M., Rozhon, W., Bucher, E., Theiler, G., and Paszkowski, J. (2009b). Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. *Genes Dev.* *23*, 939–950.
- Richards, E.J. (2009). Quantitative epigenetics: DNA sequence variation need not apply. *Genes Dev.* *23*, 1601–1605.
- Rocha, P.S.C.F., Sheikh, M., Melchiorre, R., Fagard, M., Boutet, S., Loach, R., Moffatt, B., Wagner, C., Vaucheret, H., and Furner, I. (2005). The Arabidopsis HOMOLOGY-DEPENDENT GENE SILENCING1 gene codes for an S-adenosyl-L-homocysteine hydrolase required for DNA methylation-dependent gene silencing. *Plant Cell* *17*, 404–417.
- Ronemus, M.J., Galbiati, M., Ticknor, C., Chen, J., and Dellaporta, S.L. (1996). Demethylation-induced developmental pleiotropy in Arabidopsis. *Science* *273*, 654–657.

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*, 1928–1938.

Rountree, M.R., Bachman, K.E., and Baylin, S.B. (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat. Genet.* *25*, 269–277.

Sáez-Vásquez, J., and Gadgil, O. (2010). Genome organization and function: a view from yeast and Arabidopsis. *Mol. Plant* *3*, 678–690.

Safár, J., Bartoš, J., Janda, J., Bellec, A., Kubaláková, M., Valárik, M., Pateyron, S., Weiserová, J., Tušková, R., Číhalíková, J., et al. (2004). Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. *Plant J.* *39*, 960–968.

Saintenac, C. (2009). Analyse de la distribution des crossing-overs sur le chromosome 3B du blé tendre (*Triticum aestivum*) et des facteurs influençant cette distribution. Université Blaise Pascal.

Saintenac, C., Falque, M., Martin, O.C., Paux, E., Feuillet, C., and Sourdille, P. (2009). Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (*Triticum aestivum* L.). *Genetics* *181*, 393–403.

Saintenac, C., Jiang, D., and Akhunov, E.D. (2011a). Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome Biol.* *12*, R88.

Saintenac, C., Faure, S., Remay, A., Choulet, F., Ravel, C., Paux, E., Balfourier, F., Feuillet, C., and Sourdille, P. (2011b). Variation in crossover rates across a 3-Mb contig of bread wheat (*Triticum aestivum*) reveals the presence of a meiotic recombination hotspot. *Chromosoma* *120*, 185–198.

Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegu, B., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M., and Feuillet, C. (2008a). Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. *Plant Cell Online* *20*, 11–24.

Salse, J., Bolot, S., Throude, M., Jouffe, V., Piegu, B., Quraishi, U.M., Calcagno, T., Cooke, R., Delseny, M., and Feuillet, C. (2008b). Identification and Characterization of Shared Duplications between Rice and Wheat Provide New Insight into Grass Genome Evolution. *Plant Cell Online* *20*, 11–24.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* *74*, 5463–5467.

Santos, A.P., and Shaw, P. (2004). Interphase chromosomes and the Rab1 configuration: does genome size matter? *J. Microsc.* *214*, 201–206.

Saze, H., Scheid, O.M., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.* *34*, 65–69.

- Schmidt, A., Wöhrmann, H.J.P., Raissig, M.T., Arand, J., Gheyselinck, J., Gagliardini, V., Heichinger, C., Walter, J., and Grossniklaus, U. (2013). The Polycomb group protein MEDEA and the DNA methyltransferase MET1 interact to repress autonomous endosperm development in Arabidopsis. *Plant J.* *73*, 776–787.
- Schmitz, R.J., Schultz, M.D., Urich, M.A., Nery, J.R., Pelizzola, M., Libiger, O., Alix, A., McCosh, R.B., Chen, H., Schork, N.J., et al. (2013). Patterns of population epigenomic diversity. *Nature advance online publication*.
- Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* *326*, 1112–1115.
- She, C., Liu, J., Diao, Y., Hu, Z., and Song, Y. (2007). The Distribution of Repetitive DNAs Along Chromosomes in Plants Revealed by Self-genomic in situ Hybridization. *J. Genet. Genomics* *34*, 437–448.
- Simpson, V.J., Johnson, T.E., and Hammen, R.F. (1986). *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res.* *14*, 6711–6719.
- Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C., Jackson, J.P., Kakutani, T., Koornneef, M., and Peeters, A.J. (2000). The Late Flowering Phenotype of *fwa* Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene. *Mol. Cell* *6*, 791–802.
- Steward, N., Kusano, T., and Sano, H. (2000). Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells. *Nucleic Acids Res.* *28*, 3250–3259.
- 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.
- Takuno, S., and Gaut, B.S. (2013). Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proc. Natl. Acad. Sci.* *110*, 1797–1802.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 Histone Marks and Histone Lysine Crotonylation as a New Type of Histone Modification. *Cell* *146*, 1016–1028.
- Tang, Y., Gao, X.-D., Wang, Y., Yuan, B.-F., and Feng, Y.-Q. (2012). Widespread existence of cytosine methylation in yeast DNA measured by gas chromatography/mass spectrometry. *Anal. Chem.* *84*, 7249–7255.
- 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.* *100*, 8823–8827.
- Teerawanichpan, P., Chandrasekharan, M.B., Jiang, Y., Narangajavana, J., and Hall, T.C. (2004). Characterization of two rice DNA methyltransferase genes and RNAi-mediated reactivation of a silenced transgene in rice callus. *Planta* *218*, 337–349.

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

The International Barley Genome Sequencing (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* 491, 711–716.

Trap-Gentil, M.-V., Hébrard, C., Lafon-Placette, C., Delaunay, A., Hagège, D., Joseph, C., Brignolas, F., Lefebvre, M., Barnes, S., and Maury, S. (2011). Time course and amplitude of DNA methylation in the shoot apical meristem are critical points for bolting induction in sugar beet and bolting tolerance between genotypes. *J. Exp. Bot.* 62, 2585–2597.

Trevaskis, B., Hemming, M.N., Dennis, E.S., and Peacock, W.J. (2007). The molecular basis of vernalization-induced flowering in cereals. *Trends Plant Sci.* 12, 352–357.

Vanrobays, E., Thomas, M., and Tatout, C. (2013). Heterochromatin Positioning and Nuclear Architecture. In *Annual Plant Reviews*, D.E. Evans, K. Graumann, and J.A. Bryant, eds. (John Wiley & Sons Ltd), pp. 157–190.

Varshney, R.K., Graner, A., and Sorrells, M.E. (2005). Genomics-assisted breeding for crop improvement. *Trends Plant Sci.* 10, 621–630.

Vaucheret, H. (2005). MicroRNA-Dependent Trans-Acting siRNA Production. *Sci. Signal.* 2005, pe43.

Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., Bevan, M.W., Barry, K., Lucas, S., Harmon-Smith, M., Lail, K., et al. (2010). Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* 463, 763–768.

Vongs, A., Kakutani, T., Martienssen, R.A., and Richards, E.J. (1993). *Arabidopsis thaliana* DNA methylation mutants. *Science* 260, 1926–1928.

Waterhouse, P.M., and Helliwell, C.A. (2003). Exploring plant genomes by RNA-induced gene silencing. *Nat. Rev. Genet.* 4, 29–38.

Weaver, J.R., Susiarjo, M., and Bartolomei, M.S. (2009). Imprinting and epigenetic changes in the early embryo. *Mamm. Genome* 20, 532–543.

Wen, S., Wen, N., Pang, J., Langen, G., Brew-Appiah, R.A.T., Mejias, J.H., Osorio, C., Yang, M., Gemini, R., Moehs, C.P., et al. (2012). Structural genes of wheat and barley 5-methylcytosine DNA glycosylases and their potential applications for human health. *Proc. Natl. Acad. Sci.* 109, 20543–20548.

Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T.A., Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C. (2011). High-Resolution Analysis of Parent-of-Origin Allelic Expression in the *Arabidopsis* Endosperm. *PLoS Genet* 7, e1002126.

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.

Lev-Yadun, S., Gopher, A., and Abbo, S. (2000). The Cradle of Agriculture. *Science* 288, 1602–1603.

- Yamauchi, T., Moritoh, S., Johzuka-Hisatomi, Y., Ono, A., Terada, R., Nakamura, I., and Iida, S. (2008a). Alternative splicing of the rice OsMET1 genes encoding maintenance DNA methyltransferase. *J. Plant Physiol.* *165*, 1774–1782.
- Yamauchi, T., Moritoh, S., Johzuka-Hisatomi, Y., Ono, A., Terada, R., Nakamura, I., and Iida, S. (2008b). Alternative splicing of the rice OsMET1 genes encoding maintenance DNA methyltransferase. *J. Plant Physiol.* *165*, 1774–1782.
- Yamauchi, T., Johzuka-Hisatomi, Y., Fukada-Tanaka, S., Terada, R., Nakamura, I., and Iida, S. (2009). Homologous recombination-mediated knock-in targeting of the MET1a gene for a maintenance DNA methyltransferase reproducibly reveals dosage-dependent spatiotemporal gene expression in rice. *Plant J.* *60*, 386–396.
- Yamauchi, T., Johzuka-Hisatomi, Y., Terada, R., Nakamura, I., and Iida, S. (2014). The MET1b gene encoding a maintenance DNA methyltransferase is indispensable for normal development in rice. *Plant Mol. Biol.* 1–14.
- Yan, H., Kikuchi, S., Neumann, P., Zhang, W., Wu, Y., Chen, F., and Jiang, J. (2010). Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice. *Plant J.* *63*, 353–365.
- Yang, L., and Gaut, B.S. (2011). Factors that Contribute to Variation in Evolutionary Rate among Arabidopsis Genes. *Mol. Biol. Evol.* *28*, 2359–2369.
- Yelina, N.E., Choi, K., Chelysheva, L., Macaulay, M., de Snoo, B., Wijnker, E., Miller, N., Drouaud, J., Grelon, M., Copenhaver, G.P., et al. (2012). Epigenetic Remodeling of Meiotic Crossover Frequency in Arabidopsis thaliana DNA Methyltransferase Mutants. *PLoS Genet.* *8*, e1002844.
- Zadoks, J.C., Chang, T.T., and Konzak, C.F. (1974). A decimal code for the growth stages of cereals. *Weed Res.* *14*, 415–421.
- Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell* *153*, 193–205.
- Zhang, H., and Zhu, J.-K. (2011). RNA-directed DNA methylation. *Curr. Opin. Plant Biol.* *14*, 142–147.
- Zhang, H., Chaudhury, A., and Wu, X. (2013). Imprinting in Plants and Its Underlying Mechanisms. *J. Genet. Genomics* *40*, 239–247.
- 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*, 61–69.

Annexes

Mise au point de la méthode dite de « conversion au bisulfite » sur l'ADN génomique de blé tendre

Délivrable 1 : Mise au point de la méthode dite de « conversion au Bisulfite » sur de l'ADN génomique de blé hexaploïde.

Sommaire

1-	Introduction.....	323
2-	Extraction de l'ADN génomique	325
3-	Quantification de l'ADN génomique de blé tendre.....	325
4-	Conversion au bisulfite.....	327
5-	Amplification de l'ADN-BS.....	329
6-	Analyse de la conversion au BS par séquençage.....	333
7-	Bilan.....	337

1- Introduction

Dans le cadre de l'étude des régulations épigénétiques, la méthylation des résidus cytosine de l'ADN est l'une des modifications épigénétiques de choix permettant d'évaluer l'état d'activation/ répression d'une séquence. Une telle étude requiert une technique particulière dite de « conversion au Bisulfite » (BC), un traitement chimique de l'ADN qui convertit les cytosines non méthylées en thymines alors que les cytosines méthylées restent inchangées (Aichinger & Köhler 2010). La comparaison des séquences avant / après BC donne donc accès au profil de méthylation d'une séquence d'intérêt. Chez les plantes, la méthylation est retrouvée dans trois contextes différents : CG, CHG et CHH.

Nous proposons d'adapter ces outils à l'utilisation du BC chez le blé, une espèce polyploïde. Le séquençage issu du BC devra intégrer la possibilité d'utiliser plusieurs méthodes de séquençage : la technique de Sanger (Sanger et al. 1977) et la technique de pyroséquençage 454 (Margulies et al. 2005) voir la technique Solexa (Cokus et al. 2008).

Le traitement au bisulfite de l'ADN permettra une analyse des profils de méthylation « molécule par molécule ». En effet sur le schéma ci-dessous, si l'analyse est faite dans sa globalité chaque position apparaît méthylée à 50% mais ceci peut se traduire par deux situations biologiques distinctes.

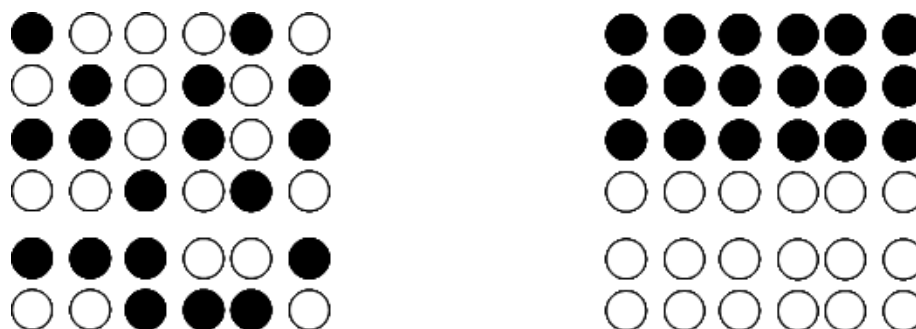


Figure 1 : Exemples de profils de méthylation d'une séquence d'ADN.

Lorsque la méthylation est considérée molécule par molécule, on constate que dans un cas (à gauche) cela est dû à une distribution mosaïque et les molécules sont homogènes entre elle. Sur le cas de droite on a toujours une méthylation globale à 50%/ par position mais ici due à des populations hétérogènes de molécules (complètement méthylées et pas méthylées du tout). D'un point de vue biologique cela peut illustrer un matériel de départ avec au moins 2 types cellulaires différents et des patrons d'expression différents.

Le séquençage bisulfite permettra cette analyse molécule par molécule avec plus ou moins de profondeur en fonction des techniques de séquençage utilisées. Ceci permettra en outre d'identifier des haplotypes de méthylation (associations de position méthylées/non méthylées) pour la région d'ADN étudiée.

2- Extraction de l'ADN génomique

L'ADNg est extrait par le kit DNeasy plant maxi kit (<http://www.qiagen.com/products/genomicdnastabilizationpurification/dneasyplantsystem/dneasyplantminikit.aspx>). L'extraction de l'ADNg est réalisée à partir de plantules au stade 3-4 feuilles poussées en chambre de culture Sanyo sur vermiculite. Environ 5 plantules sont poolées pour obtenir une quantité d'environ 1g de matériel. Après broyage manuel dans l'azote liquide, les cellules sont lysées à 65°C en présence de RNase et précipitées par centrifugation. Le surnageant est ensuite passé à travers une colonne afin de retenir l'ADN. Celui-ci est lavé avec d'être élué dans de l'eau milliQ. L'ADNg est aliquoté et conservé à -20°C ; un aliquot ne sera décongelé qu'une fois puis conservé à 4°C. Le kit préconise de faire 2 éluations successives appelées « éluat 1 » et « éluat 2 » qui ont été gardées indépendamment dans les étapes suivantes.

3- Quantification de l'ADN génomique de blé tendre

L'ADNg est dosé par deux méthodes :

- Nanodrop (<http://www.nanodrop.com/>)

- Sur gel en référence à une gamme étalon d'ADN du phage λ (DNA quantification standards phage λ DNA 50 assays, Gibco BRL, Life Technologies)

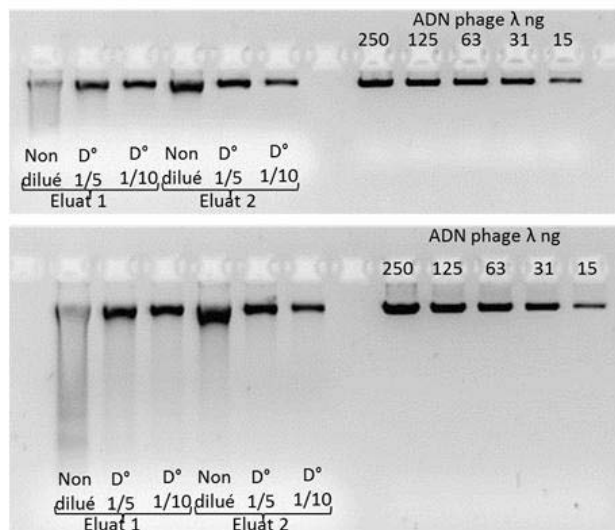


Figure 2 : Quantification sur gel d'une extraction d'ADNg de blé

L'ADNg utilisé dans cette expérience a été dosé au nanodrop 84,5ng/ μ l pour l'éluat 1 et 25,2ng/ μ l pour l'éluat 2. Il est ensuite déposé sur gel d'agarose à 1,5% additionné de 0,5 μ g/ml de BET. Les dépôts correspondent à 10 μ l des deux éluats non dilués, puis dilués respectivement au 1/5^{ème} et au 1/10^{ème}. La gamme étalon d'ADN de phage λ est comprise entre 15 et 250ng d'ADN).

A partir de ce gel, la mesure de l'intensité de signal des bandes a été estimée avec le Versadoc™ Imaging System (Biorad) et quantifié à l'aide du logiciel Quantity One® (Biorad). La concentration en ADNg de l'éluat 1 est estimée à 68,2 \pm 8,1ng/ μ l et celle de l'éluat 2 à 18,5 \pm 2,4 ng/ μ l. L'écart type des concentrations calculées est de 12% environ dans les deux cas la concentration est surévaluée lors d'un dosage au nanodrop d'un facteur 1,2 pour l'éluat 1 et 1,4 pour l'éluat 2.

4- Conversion au bisulfite

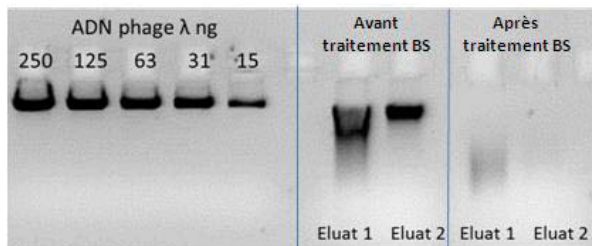
La conversion au bisulfite est réalisée avec un kit Zymo Research : EZ DNA Methylation-Gold™ Kit. La commande se fait via Proteigene

http://catalogue.proteigene.com/epigenetique/1-d5005-d5006-ez-dna-methylation-gold-kit.html?cPath=505_623_497_503

130 μ l de réactif de conversion est rajouté à 200ng et 500ng d'ADNg contenus dans 20 μ l. L'ADN est dénaturé et convertit lors d'une même étape qui se déroule en thermocycler selon le programme suivant : 98°C pendant 10min, 64°C pendant 2,5 heures et 4°C jusqu'à poursuite de la conversion. La conversion de l'ADN se poursuit ensuite sur colonne par l'étape de desulphonation et de lavage de l'ADN (ensuite appelé ADN-BS). L'ADN-BS est élué dans un volume de 10 μ l / colonne et peut être conservé jusqu'à 20h à 4°C pour une

utilisation immédiate ou conservé à -20°C pour une utilisation ultérieure, sous forme d'aliquots pour éviter des étapes de congélation-décongélation. Un stockage sur une plus longue durée se fera à -80°C .

5 réactions sont menées en parallèle à partir d'ADNg provenant de l'éluat 1 (~500ng d'ADNg initial/réaction) et de l'éluat 2 (~200ng d'ADNg initial/réaction). Les 5 réactions (10 μl d'éluat / réaction) sont ensuite regroupées ($V_T=50\mu\text{l}$). L'ADN-BS est vérifié sur gel par comparaison à l'ADNg initial.



10 μl d'ADNg initial dilué au 1/5 et 10 μl d'ADN-BS pur sont déposés sur un gel d'agarose à 1% additionné de 0.5 μg BET/ml.

Figure 3 : Vérification sur gel du traitement au BS

L'ADN-BS forme un smear prouvant que le traitement au bisulfite dégrade en partie l'ADNg d'où l'importance de ne pas partir d'une quantité trop faible d'ADNg au départ (rester autour de 0.5 μg).

Ceci implique que les éluat 1 et 2 issus du kit DNAeasy ne doivent pas être poolés et que seul l'ADNg issu de l'éluat 1 est assez concentré pour être utilisé dans la conversion au Bisulfite. Pour le futur seul l'éluat 1 sera conservé.

A partir de ce stade, seul l'ADN-BS issu de l'éluat 1 a été considéré pour la suite du protocole

5- Amplification de l'ADN-BS

L'amplification sur ADN-BS nécessite des amorces qui seront dégénérées, c'est-à-dire où les cytosines sont remplacées par les bases dégénérées, R ou Y selon qu'il s'agisse de l'amorce forward ou reverse. Il s'agit également d'éviter les zones contenant des cytosines susceptibles d'être converties (éviter absolument les dinucléotides CG qui sont les principaux sites de méthylation). Il existe plusieurs méthodes possibles pour définir les amorces utilisables en séquençage bisulfite. Ces méthodes seront détaillées dans un autre mémo technique.

Dans une étude pilote, nous avons utilisé 3 amplicons du gène *VRN1* fournis par l'équipe de Jérôme ENJALBERT de l'INRA du Moulon : VRN1-1, VRN1-2 et VRN1-3. Nous avons

convenu que la séquence des amorces sera mise à disposition de Biogemma lorsque le travail de J. ENJALBERT sera publié (la publication est en cours de soumission).

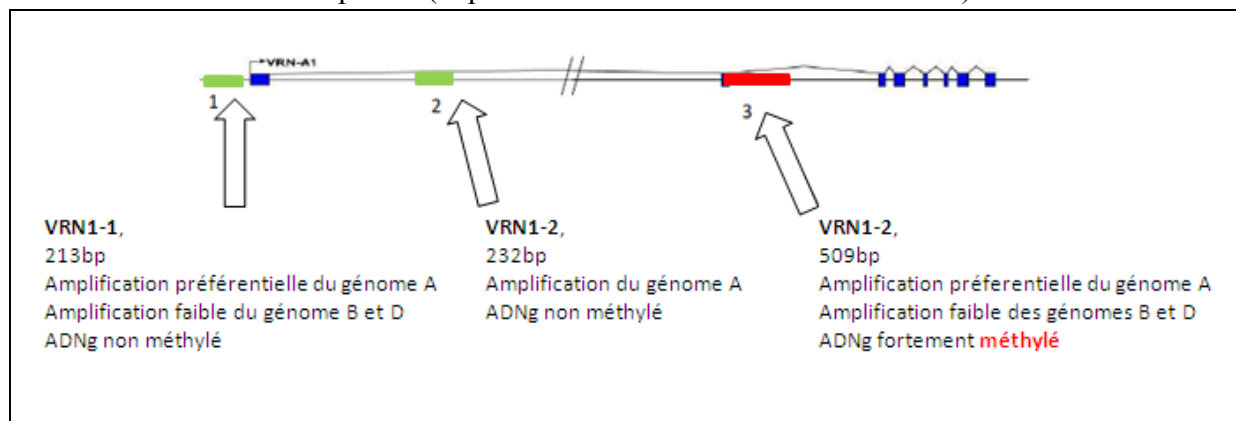


Figure 4 : le gène VRN1 de blé est composé de 8 exons (en bleu) et de 7 introns (traits). Les trois amplicons sont localisés dans des régions dépourvues de méthylation (VRN1-1 et VRN1-2) ou fortement méthylé (VRN1-3).

Pour ces 3 amplicons, les amorces ont été construites en évitant la présence de CG et de CHG. Comme il s'agit d'une séquence codante, les sites CHH seront faiblement ou non méthylés donc convertis. Seul un brin de l'ADN-BS a été étudié dans cette expérience. Les amplifications PCR sur de l'ADN-BS nécessitent l'utilisation de Taq Polymérase de type Hot Start : [AmpliTaq Gold® 360 Master Mix](#) (Applied biosystem). Les amplifications sont réalisées dans un volume de 40µl final, 0,5µM concentration finale de chaque amorce, 20µl de Master Mix (recommandée par le fournisseur), 1µg d'ADN-BS et de l'eau milliQ qsp 40µl. Il est possible de réduire la quantité d'ADN-BS dans la PCR (<1µg/réaction recommandé par le fournisseur).

Le programme sur thermocycler suit le modèle suivant :

Activation de l'AmpliTaq Gold® 360 Master Mix	5 à 10 min à 95°C
10 cycles en touch down	1min 95°C
	1min 65°C à 55°C (1°C/cycle)
	1min30 72°C
10 cycles en touch down	1min 95°C
	1min 55°C à 50°C (0,5°C/cycle)
	1min30 72°C
20 cycles	45sec 95°C
	45sec 50°C
	1min 72°C
	7min 72°C
	Pause 4°C

L'ADN amplifié est vérifié sur gel

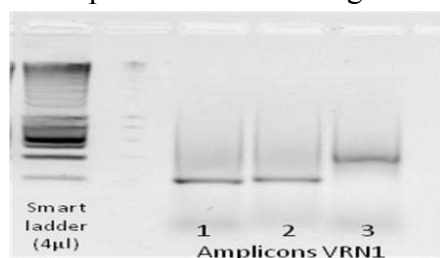


Figure 6 : Vérification des produits d'amplification

Une migration sur gel à 1% d'agarose a été faite avec 6/40µl des produits d'amplification.

La quantité de chaque produit d'amplification est de l'ordre de 100-150µg / 40µl. 1/10^{ème} de ces produits d'amplification a été utilisé pour un séquençage de type Sanger et le reste pour un séquençage selon la méthode du 454 (Roche).

6- Analyse de la conversion au BS par séquençage

La vérification s'est faite par séquençage classique auprès d'un prestataire de service (méthode de Sanger). La ligation des produits PCR se fait en vecteur pGEMT-easy (Promega). Les ligations sont réalisées dans un volume final de 10µl en présence de 3µl du produit PCR pendant une nuit à 4°C. Les clones positifs sont repiqués sur une plaque 96 fournie par le prestataire de séquençage. Les plasmides sont séquencés avec les amorces SP6 et T7. Les résultats du séquençage Sanger sont analysés avec l'outil Kismeth (<http://katahdin.mssm.edu/kismeth/revpage.pl>). Un logiciel utilisable en interne par Biogemma et adapté au génome polyploïde est en cours de développement par le GReD et Biogemma.

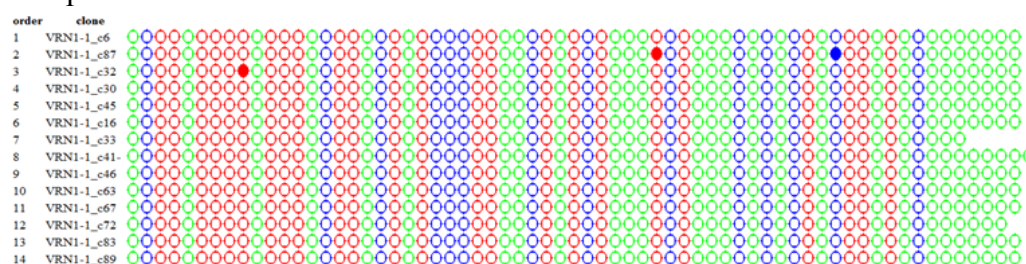
- VRN1-1

72 clones VRN1-1 ont été séquencés. Les séquences obtenues se séparent en trois groupes, correspondant aux trois génomes confirmant que les amorces pour l'amplicon VRN1-1 ne sont pas spécifiques d'un génome. Globalement, les amplicons des trois génomes sont peu ou pas méthylé (tableau ci-dessous).

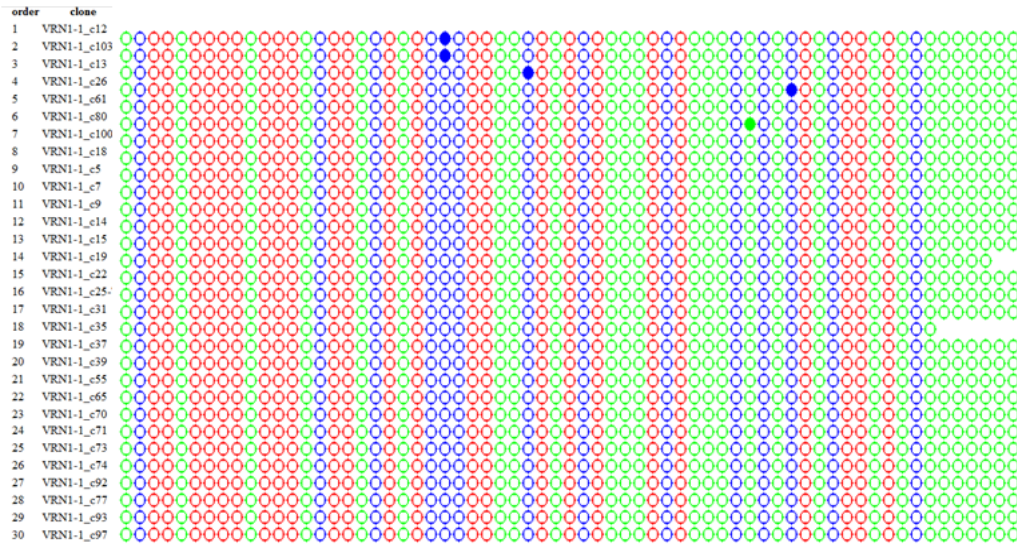
	Groupe 1		Groupe 2		Groupe 3	
	Total méthylé	Total non méthylé	Total méthylé	Total non méthylé	Total méthylé	Total non méthylé
CG	0,68%	99,31%	0%	100%	0,54%	99,45%
CHG	0,51%	99,48%	0,95%	99,04%	0,85%	99,14%
CHH	0%	100%	0,12%	99,87%	0,59%	99,40%
all	0,352%	99,647%	0,257%	99,742%	0,634%	99,365%

Les profils de méthylation sont visualisés par la sortie graphique de Kismeth. Chaque rond correspond à une cytosine non méthylées quand ils sont vides et méthylées quand ils sont pleins. Les différentes couleurs correspondent aux différents contextes de cytosine (CG en rouge, CHG en bleu et CHH en vert).

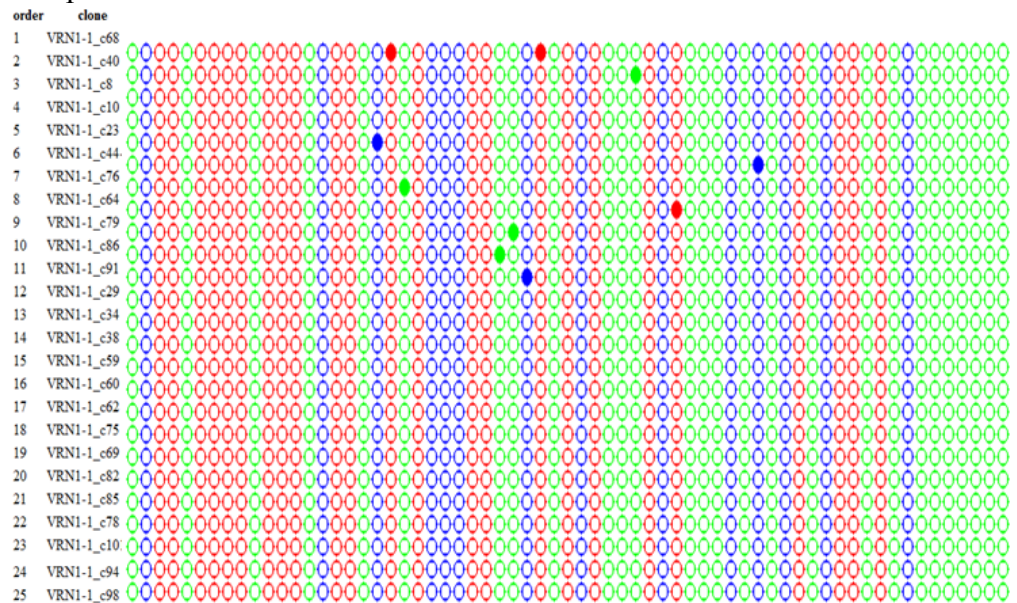
Groupe 1



Groupe 2



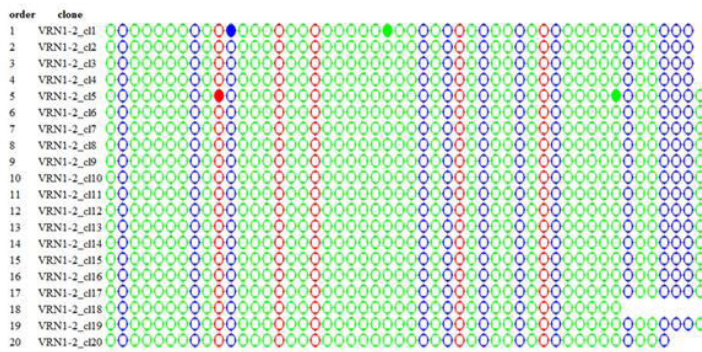
Groupe 3



- VRN1-2

20 clones de VRN1-2 ont été séquencés. Un seul génome est ciblé (génome A). Globalement l'ADN est non méthylé (voir tableau).

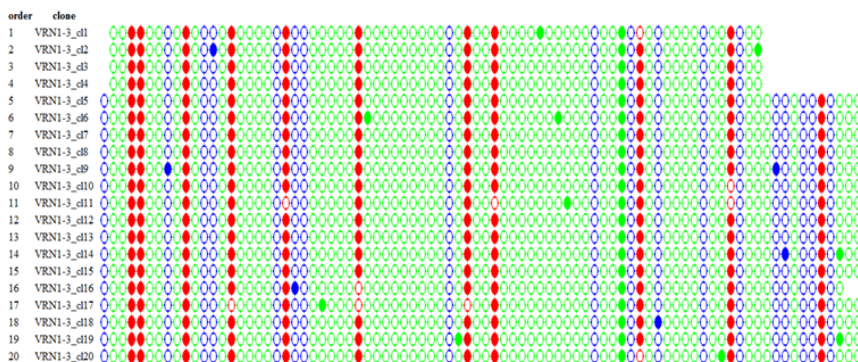
	Total méthylé	Total non méthylé
CG	1%	99%
CHG	0,42%	99,57%
CHH	0,30%	99,69%
all	0,405%	99,594%



- **VRN1-3**

20 clones de VRN1-3 ont été séquencés. Un seul génome est ciblé (génome A). Globalement l'ADN est fortement méthylé mais uniquement au niveau des sites CG (voir tableau).

	Total méthylé	Total non méthylé
CG	95,81%	4,18%
CHG	1,78%	98,21%
CHH	2,78%	97,21%
All	14,864%	85,135%



7- Bilan

Nos résultats sont en adéquation avec les résultats attendus. Le protocole décrit peut-être appliqué à n'importe quel type de séquence dont on veut connaître le profil de méthylation ADN.

Aichinger, E., and Köhler, C. (2010). Bisulphite Sequencing of Plant Genomic DNA. In *Plant Developmental Biology*, L. Hennig, and C. Köhler, eds. (Totowa, NJ: Humana Press), pp. 433–443.

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, 215–219.

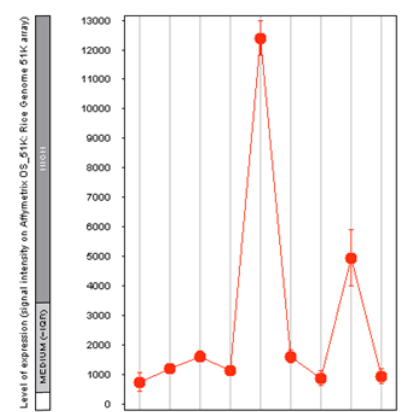
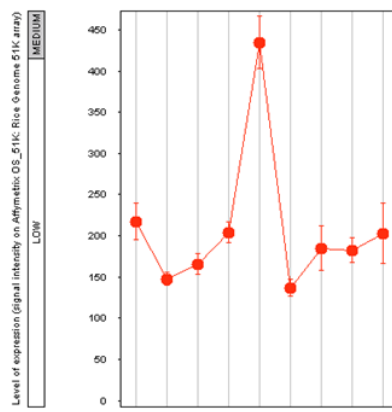
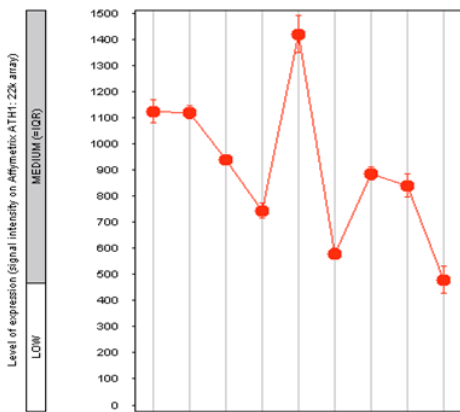
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74, 5463–5467.

Arabidopsis
AT5G49160

Riz
OsMET1a : Os03g58400

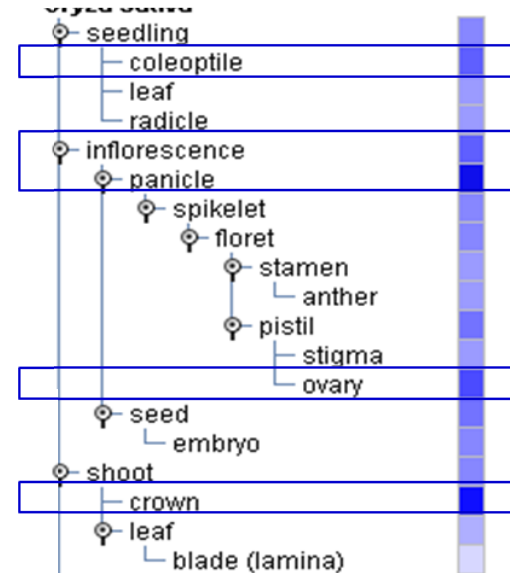
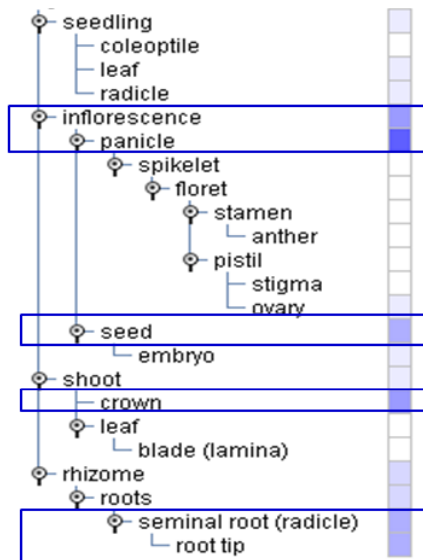
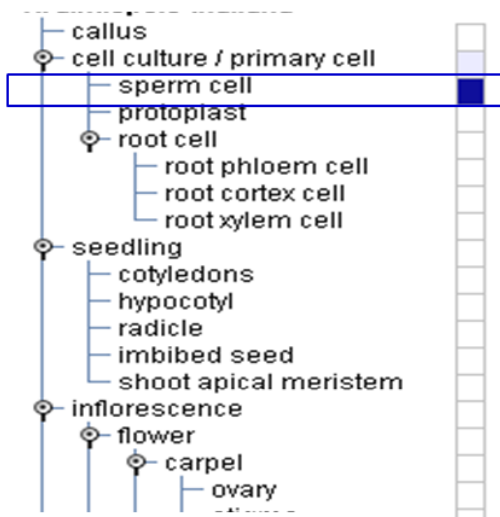
OsMET1b : Os07g08500



Stage of development:

Stage of development:

Stage of development:



Profil d'expression des gènes *MET1* chez *Arabidopsis* et le riz

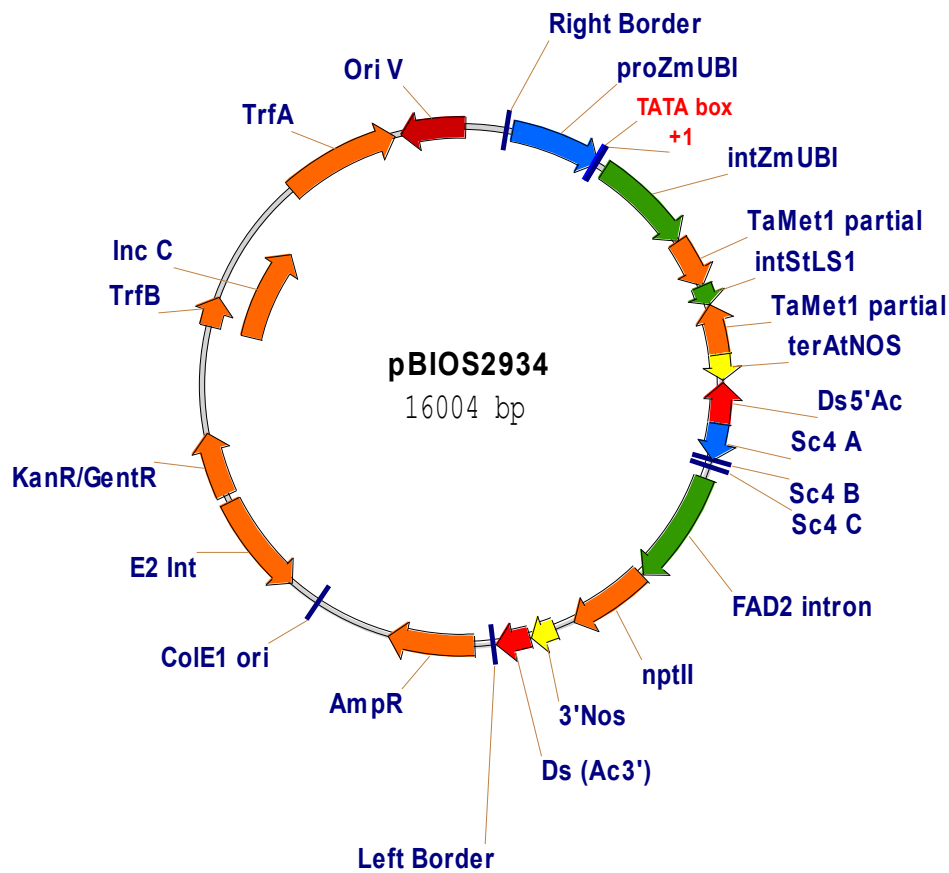
Source : Genevestigator, www.genevestigator.com

Les courbes rouges correspondent aux profils d'expression des gènes *MET1* selon différents stades de développement. La barre sur la gauche précise s'il s'agit d'un gène avec une expression faible (blanc), moyenne (gris clair) ou forte (gris foncé).

Les graphiques du dessous précisent dans quels tissus l'expression est la plus forte, parmi l'ensemble des tissus étudiés. Plus la couleur bleue est foncée, plus l'expression du gène est forte.

noms	A1	A2	amorces C (C1 ou C2)	SNP1	SNP2
Met1_tilling_pos241_CAG/TAG	GAAGGTGACCAAGTTCATGCTATGAACAACCTTTGTATGCTTCTCTG	GAAGGTCGGAGTCAACGGGATTCATGAACAACCTTTGTATGCTTCTCTA	TAGTCTTTGGCCACTGGCCTGTT	C	T
Met1_tilling_pos501_CAG/TAG	GAAGGTGACCAAGTTCATGCTGAGGATGCTGAGGAAGAGGTAC	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	GTTGATGGTGGAGTCTTTGGCAACCTT	C	T
Met1_tilling_pos549_CGA/TGA	GAAGGTGACCAAGTTCATGCTACTCCACCATCAACACGGTCTC	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	ACCTTCCATCTCATTTCTTTGCTAGTTT	C	T
Met1_tilling_pos565_CAA/TAA	GAAGGTGACCAAGTTCATGCTTCCATCTCATTTCTTTGCTAGTTTG	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	CAAGGTTGCAAAAGACTCCACCATCAA	C	T
Met1_tilling_pos737_CAG/TAG	GAAGGTGACCAAGTTCATGCTGTCATATGCTTTGTTGAGTACATGTTTC	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	CCATGAACCAATTTTGGCAACCATCAGATTT	C	T
Met1_tilling_pos908_TGG/TGA	GAAGGTGACCAAGTTCATGCTGTCATCTCCAATCAATGCCCTTGA	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	CTCAGCTTCAGAAATCTCTTTTCTTACTT	A	G
Met1_tilling_pos1022_TGG/TGA	GAAGGTGACCAAGTTCATGCTAAGAAGGCCACCTTCTCAGGCT	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	AAAGGCTGCCCCATGGAAATTTATGTA	A	G
Met1_tilling_pos1214_CAA/TAA	GAAGGTGACCAAGTTCATGCTGAAGAAGGCCACCTTCTCAGGT	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	AAAGGCTGCCCCATGGAAATTTATGTA	A	G
Met1_tilling_pos2150_CAA/TAA	GAAGGTGACCAAGTTCATGCTGCTCCAGCTGGTGAAGCATTTAGGC	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	GGTAACCTATAGTGTCCATGACTCTTAT	C	T
Met1_tilling_pos2348_CAA/TAA	GAAGGTGACCAAGTTCATGCTCTCCAGCAAGTTTAGCTGCTTG	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	GCAAGTTATCCACAACACACAGCAGCTT	C	T
Met1_tilling_pos2554_CAA/TAA	GAAGGTGACCAAGTTCATGCTACTGAACCTTTACTCCATGGGCTTTG	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	ATTCTGATGACTGTGCTCAACTTCTGAA	C	T
Met1_tilling_pos2563_TGG/TAG	GAAGGTGACCAAGTTCATGCTATGAACAGATTCAAACCAAGGCCCATG	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	CAGGCTTTTCTGGCCATGAACAGAT	C	T
Met1_tilling_pos2563_TGG/TAG	GAAGGTGACCAAGTTCATGCTGAACAGATTCAAACCAAGGCCCATG	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	TCCGCAAAAAGAGGAAATGCTAAAATCAT	A	G
Met1_tilling_pos2563_TGG/TGA	GAAGGTGACCAAGTTCATGCTGAACAGATTCAAACCAAGGCCCATG	GAAGGTCGGAGTCAACGGGATTCATGAGGATGCTGAGGAAGAGGTAT	TCCGCAAAAAGAGGAAATGCTAAAATCAT	A	G

Liste des amorces utilisées pour le premier crible des populations de TILLING



Représentation schématique du vecteur utilisé pour la deuxième vague de RNAi contre *MET1*

La séquence de 630pb ciblant *MET1* a été clonée afin d'obtenir deux séquences en répétition inverse. Les deux séquences sont séparées par l'intron du gène *StSL1* et sous contrôle du promoteur du gène de l'ubiquitine du maïs (*proZmUbi*). L'arrêt de la transcription est assuré par le terminateur NOS (*terAtNOS*). La résistance kanamycine (*KanR*) permet la sélection bactérienne et le gène *NptII* la sélection en plante. Les séquences *Right* et *Left Border* délimitent la portion du vecteur qui sera intégrée au génome de la plante.

Gène	Genome	Nom amplicon	Amorces sens		Amorces anti-sens		Taille amplicon
			Nom	Séquence 5'3'	Nom	Séquence 5'3'	
Ppd1	D		MeT_Ppd1-D1_pos395-fwd	ACGCCTCCCACTACACTG	MeT_D1-Ppd1_pos1303-rev	CCCAGTACCAGTAAAAGATGC	908
	A	a	MeT_Ppd1-A1_pos7309-fwd	AGGTTTCGATTGTTATTACAG	MeT_A1-Ppd1_pos8270-rev	CACGTTGACCCAGAAAAGG	961
	B	b	MeT_Ppd1-B1_pos10147-fwd	AGGTTTCGATTATTACTCACGG	MeT_B1-Ppd1_pos11050-rev	CCAGTACCATACCAGTAAAAG	903
	D	c	MeT_Ppd1-D1_pos1303-fwd	GCATCTTTACTGGTACTGGG	MeT_D1-Ppd1_pos2227-rev	GCTGGAAAAGAAAAATGTTGG	924
	A	d	MeT_Ppd1-A1_pos8270-fwd	CCTTTTCTGGGTCAACGTG	MeT_A1-Ppd1_pos9134-rev	AACGTGTTGTAGTGACTTGAGC	864
	B	e	MeT_Ppd1-B1_pos11050-fwd	CTTTTACTGGTATGGTACTGG	MeT_B1-Ppd1_pos11935-rev	CATCACCTGGAAAACATATTG	885
	D	f	MeT_Ppd1-D1_pos2227-fwd	CCAACATTTTTTCTTTCCAGC	MeT_D1-Ppd1_pos3138-rev	CGATGATCCACCATTCTGTG	911
	A	g	MeT_Ppd1-A1_pos10347-fwd	TGTTGTAGTGAGTGCTCAACC	MeT_A1-Ppd1_pos11362-rev	CGATGGATTGACCAAACCTGG	1015
	B	h	MeT_Ppd1-B1_pos11935-fwd	CAATATGTTTTCCAGGTGATG	MeT_B1-Ppd1_pos12862-rev	GGATTGATCACGTTGGACTG	927
	D	i	MeT_Ppd1-D1_pos3138-fwd	CACGAATGGTGGATCATCG	MeT_A1D1-Ppd1_pos3604-rev	GTATAGCTAGGTGTGTGGC	466
	A	j	MeT_Ppd1-A1_pos11362-fwd	CCAGTTTGGTCAATCCATCG	MeT_B1-Ppd1_pos13421-rev	ATGAGACGAGATGAGGACC	559
	B	k	MeT_Ppd1-B1_pos12862-fwd	CAGTCCAACGTGATCAATCC			
	vrn2	A		MeT_Vrn2-A_pos1098-fwd	AGAAACGGAGGGAGTACATC	MeT_Vrn2-AD_pos1979-rev	
D			MeT_Vrn2-D_pos8398-fwd	ATGATTGATCAAACGATGGAC		CCAGAAAAGCTTACGCTGAG	932
A			MeT_Vrn2-AD_pos1979-fwd		MeT_Vrn2-A_pos2936-rev	TCAAGCAATACATATGCGCTC	957
D				CTCAGCGTAAGCTTTTCTGG	MeT_Vrn2-D_pos10285-rev	AAGCAATACATATGCTCAATTCTC	955
A			MeT_Vrn2-A_pos2936-fwd	TGAGCGCATATGTATTGCTTG	MeT_Vrn2-AD_pos3422-rev		486
D			MeT_Vrn2-D_pos10350-fwd	GCAGCTTGATTAAGTCTTGC		AGAGCCCATCGTAAGTAG	562
FTA	A		MeT_FTADh1_pos3-fwd		MeT_FTADh1_pos774-rev		771
	D			AGGGTGACCTTCGGGAAC		GAAGAGCACGAGCACGAAG	764
Rht	A		MeT_Rht-A_pos1-fwd	ACCTATTGTTGATCCTCGGA	MeT_Rht-A_pos948-rev	GTCATGTCGTCTTCCAATTC	947
	B		MeT_Rht-BD_pos1-fwd		MeT_Rht-BD_pos768-rev		767
	D			ATGAAGCGGGAGTACCAGG		CAGCAAGGGTATCTGCTTC	770
	A		MeT_Rht-A_pos948-fwd	GAATTGGAAGACGACATGAC	MeT_Rht-A_pos1909-rev	GTCAAGGGCCTTTTGATTC	961
	B		MeT_Rht-BD_pos768-fwd		MeT_Rht-BD_pos1830-rev		1062
	D			GAAGCAGATACCCTTGCTG		CATGCGGAGGTGGCGATC	1065
	A		MeT_Rht-A_pos1909-fwd	GAATCAAAGAGGCCCTTGAC	MeT_Rht-A_pos2814-rev	GTGAGCTGTATTTGAGGTTTG	905
	A		MeT_Rht-A_pos2814-fwd	CAAACCTCAAATACAGCTCAC	MeT_Rht-A_pos3508-rev	GAGGTGGACGAGCTGCTG	694
Vrn3	B		MeT-Vrn3-B_pos1-fwd	TGTGTGGCATGTAICTATAAG	MeT-Vrn3-B_pos916-rev	CTGGTTGTGTGCGTGCTG	916
	B		MeT-Vrn3-B_pos916-fwd	GACCACGCACACAACCAG	MeT-Vrn3-B_pos1969-rev	CAAACATGTGAACAGAAGCAT	1053
	B		MeT-Vrn3-B_pos1969-fwd	ATGCTTCTGTTACATGTTTG	MeT-Vrn3-B_pos2741-rev	GAAGAGTACGAGCACGAAG	772
	B		MeT-Vrn3-B_pos2741-fwd	CTTCGTGCTCGTACTCTTC	MeT-Vrn3-B_pos3525-rev	ATTGATGAGCGCACAGAC	784

Amorces utilisées pour le crible de la méthylation des gènes cibles avec les enzymes sensibles à la méthylation

Les noms d'amplicons de a à k correspondent à ceux présentés dans la Fig 36.

gène	génotypes	fwd pour PCR1	rev pour PCR1	n° amplicon
Ppd-1	AB	ACACGACGCTCTCCGATCTATYTTTGATGTGTA AAAAATTTYAG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTTRATATACACATRTATTACAT	1
	AB	ACACGACGCTCTCCGATCTAGAAAAGATATGYGGATYYAG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTRRATRATTCRATCCCAATC	2
	ABD	ACACGACGCTCTCCGATCTYTTGGTTTTYAGTGATGYGT	GACTGGAGTTCAGACGTGTGCTCTCCGATCTRTCTRAATRRCACTTCCACT	3
	ABD	ACACGACGCTCTCCGATCTATGGAYGGTGTGAGTATYTG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTRATATTATRCTTTRCACCTC	4
Vrn-1	AB	ACACGACGCTCTCCGATCTTYAATGTTGTTAAAGTYTYTTTG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTATCATTAAATCARTACCTTCCT	5
	D		GACTGGAGTTCAGACGTGTGCTCTCCGATCTCATTAAARTCARTACCTTCCT	6
Vrn-2	AD	ACACGACGCTCTCCGATCTGATYTYGTGAYYATTTTAAATAG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTRCCCATATCCCAARCA	7
	AD	ACACGACGCTCTCCGATCTTGTGGTYAAATAAAATGAAATAAG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTAATTRTTTCATCTTTRCCRAT	8
Rht-1	BD	ACACGACGCTCTCCGATCTATGGGGATGGGYGGYGTG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTRCCTCCACCACARACTCC	9
	BD	ACACGACGCTCTCCGATCTYAAGYAGGGGATGYAGTGG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTRCCTCTCRCTCCACCAC	10
	BD	ACACGACGCTCTCCGATCTYTGTTGGTGGAGYAGGAGG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTCTCTCTCCACCCTTRTA	11
Rht-A1 (promoteur)	A	ACACGACGCTCTCCGATCTAYGGTTTTGGTTGGTTGYAT	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTRCCTCCATCTTTATACCA	12
	A	ACACGACGCTCTCCGATCTTTTTAAATATTTTTGYAAGGTAG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTACCTCRCTARCTTATAAC	13
	A	ACACGACGCTCTCCGATCTGYAGYAAYGATTTTTATTATTG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTCCRCRTACAACCACAAC	14
Vrn-3 FT	AB	ACACGACGCTCTCCGATCTTGTGGTGYTYGTGTTAAG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTCCCTATATCTTTCTTTRRATT	15
	AB	ACACGACGCTCTCCGATCTATATATYTYGGGAAAATGAT	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTCATRATCACCACRAAATAAT	16
	ABD	ACACGACGCTCTCCGATCTAGGYGAYGAGATYYGTG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTCCTCATCTTRCCRCC	17
	B	ACACGACGCTCTCCGATCTGAATAAATGATYTYTAGYTTTTG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTRRATRTACAAATTTARTACTACCA	18
TaMET-1 (groupe 5)	AD	ACACGACGCTCTCCGATCTGGAAYATTGATTTTTYATGYTT	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTCAARCTRCACACAT	19
	B	ACACGACGCTCTCCGATCTTYYATGYAAGGAAGTGGYAG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTAATCCTTTACCTTTTRTCATC	20
	D	ACACGACGCTCTCCGATCTTYYATGYAAGGAAGTGGYAG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTCTTTCCCTTTTRTRTCAC	21
	B	ACACGACGCTCTCCGATCTGTTTTTYGGGAGGGATTG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTRCCATCATCATTACCC	22
	B	ACACGACGCTCTCCGATCTGGTGGGTTTYAAYTTTTGAT	GACTGGAGTTCAGACGTGTGCTCTCCGATCTAACACATRTCAATRTAAAACC	23
	D	ACACGACGCTCTCCGATCTGAAYTGTGAAGGAYTATYTG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTARATCCTCRACAAAACCTCAT	24
TaMET-1 (groupe 7)	ABD	ACACGACGCTCTCCGATCTATYGAAYTYTAATTAGGATGT	GACTGGAGTTCAGACGTGTGCTCTCCGATCTCCCTRTAAARITTTARAATCA	25
	D	ACACGACGCTCTCCGATCTGGGYGAGTTTTGTATAAYATA		26
	A	ACACGACGCTCTCCGATCTGGGYGAGTTTTGTATAAYTTA	GACTGGAGTTCAGACGTGTGCTCTCCGATCTAAARCAAAAACAATAARAARAC	27
	B	ACACGACGCTCTCCGATCTGGGYGAGTTTTGTATAAYTTA		28
TaMET-1 (groupe 2)	D	ACACGACGCTCTCCGATCTYTGTTGGTGGYAGAGAAGAATG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTACACARACTCCTTAAATTCAC	29
	BD	ACACGACGCTCTCCGATCTTGATTTAYTTGGAGTGGATGG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTTTCAACTRTTTTTCTCTTCC	30
Vrn-1	A	ACACGACGCTCTCCGATCTGYTTTAYTTGGAGTGGATGAA	GACTGGAGTTCAGACGTGTGCTCTCCGATCTTTCACAACRTTTCTCTTCC	31
	ABD	ACACGACGCTCTCCGATCTAAATGATTTGGGAAAAGTAAATTT	GACTGGAGTTCAGACGTGTGCTCTCCGATCTAAAAAATTTTTAAAAAATCTAACCC	VRN1_1
	A	ACACGACGCTCTCCGATCTGGGGATAAGTAATTGTTATGTTTTG	GACTGGAGTTCAGACGTGTGCTCTCCGATCTATCCAAAAATCAACAACTACAT	VRN1_2
	BD	ACACGACGCTCTCCGATCTATGGATTGTTGGTYATTGTTGAT	GACTGGAGTTCAGACGTGTGCTCTCCGATCTARCCCTCAAAAACCTCARCC	VRN1_3

Amorces utilisées pour la production des amplicons pour le séquençage MiSeq

Les amorces données ici sont les amorces utilisées pour la PCR1.



Chapter 6

HETEROCHROMATIN POSITIONING AND NUCLEAR ARCHITECTURE

Emmanuel Vanrobays, Mélanie Thomas and
Christophe Tatout

GReD Laboratory, UMR CNRS 6293, INSERM U1103, University Blaise Pascal, Aubière, France

Abstract: Heitz (1928) first described the two states of chromatin known as *euchromatin* and *heterochromatin*. Heterochromatin is often described as the gene-poor part of the genome associated with a silent and condensed state of chromatin inaccessible to transcription factors. However, this simple view has been challenged many times as heterochromatin is indeed transcribed and differences between the two states of chromatin depend on many other criteria. Data collected from the model species *Arabidopsis thaliana* indicate that heterochromatin relies on the repetitiveness of specific DNA sequences such as satellites, transposable elements and ribosomal DNA (rDNA) but also on epigenetic marks specifically associated with these repeated arrays. Recent studies on small RNA pathways have highlighted the central role of the RNA-directed DNA methylation pathway in heterochromatin specification indicating that heterochromatin is indeed an epigenetic state required for many other genome functions including chromosome segregation, gene regulation or maintenance of genome stability. Heterochromatin is a specific feature of eukaryotes and preferential localization at the nuclear periphery and to the nucleolus is observed in most organisms. This spatial organization of heterochromatin is maintained through the cell cycle although DNA is replicated, chromatin is condensed into chromosomes, and the nuclear envelope is disrupted and reformed. Whether spatial positioning participates in heterochromatin function and how plant cells are able to establish and maintain such a genome organization across the cell cycle is still largely unknown in *Arabidopsis*. Mechanisms leading to the appropriate positioning of heterochromatin within the nucleus will be discussed in the light of data coming from other species. A better understanding of

heterochromatin organization and functions will be an important field of investigation for plants such as maize and wheat with considerably larger genome sizes than *Arabidopsis*.

Keywords: heterochromatin; epigenetics; chromocenters; chromatin organization; chromosome; chromosome territory; histone

6.1 Heterochromatin structure

In order to better understand how heterochromatin is assembled and what functions are assumed by it, it is first necessary to describe some structural aspects of heterochromatin. Emil Heitz (1928) historically identified heterochromatin as the nuclear material that remains highly condensed within the interphase nucleus. He named these regions 'heterochromatin' to distinguish them from the regions showing variable staining and condensation, which he called 'euchromatin'. The boundaries of heterochromatin as specified by such cytological analyses are not very well defined and may vary in diverse tissues or with different analytical techniques. Subsequently, heterochromatin was subdivided into two classes, constitutive and facultative heterochromatin, depending on its persistent presence throughout the cell cycle (Brown, 1966). Formation of facultative heterochromatin at specific genomic loci is fundamentally important in defining cellular properties such as differentiation, stress response and reaction to developmental, physiological, or environmental stimuli.

At the same time, renaturation kinetics of DNA, also known as Cot analysis, gave striking results when applied to genomic DNA from eukaryotes (Britten and Kohne, 1968). The eukaryotic genome includes three distinct fractions according to their rates of renaturation because repeated sequences renature faster than single DNA sequences (Figure 6.1a). The repeated nature of DNA turns out to be a very interesting criterion for better defining heterochromatin, as most of heterochromatic sequences fall into the highly repeated (HR) fraction whereas euchromatin sequences are included in the single or low (SL) fraction. Since these early cytological and kinetic characterizations, euchromatin and heterochromatin are subsequently distinguished according to many other criteria such as their chromosome distribution, DNA sequence composition, protein binding or epigenetic features as will be detailed below.

6.1.1 Heterochromatic sequences

In small genome species such as *Arabidopsis* (150Mbp), heterochromatin represents 10–15% of the whole genome (*Arabidopsis* Genome Initiative, 2000). The *Arabidopsis* genome is organized in five linear chromosomes with dedicated domains made of heterochromatin (Figures 6.1b and 6.1c). The

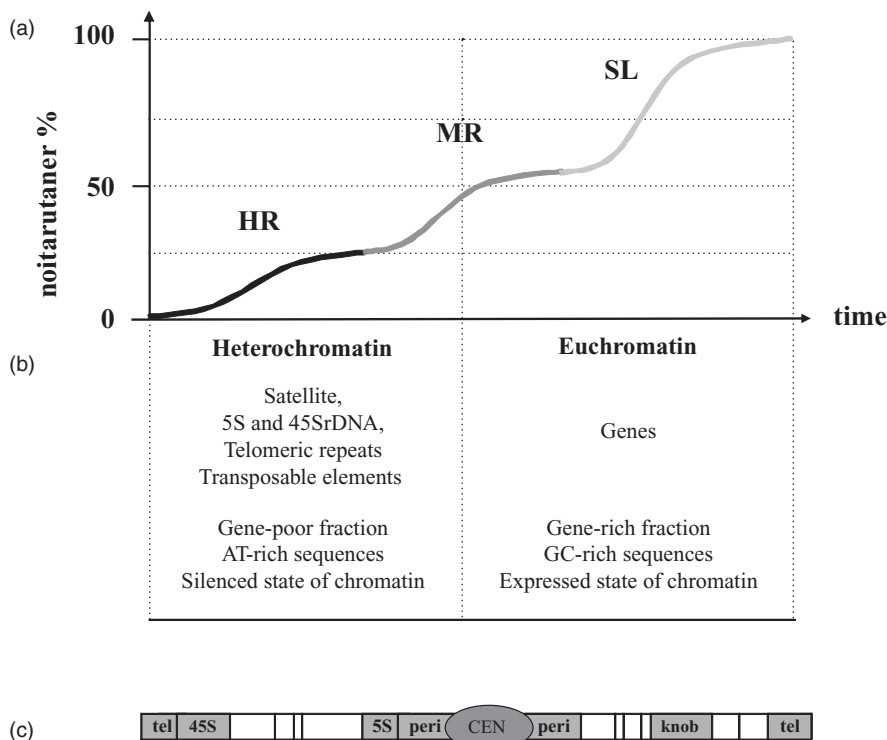


Figure 6.1 Distinctive structural features of the two states of chromatin. a) The renaturation curve of double-strand DNA highlights the three main sequence components of the eukaryotic genome in respect to their repetitiveness. The curve expressed as a percentage of renaturation (vertical axis) over time (horizontal axis) consists of highly repetitive (HR), moderately repetitive (MR), and single/low-copy (SL) components characterized by fast, intermediate and slow reassociation, respectively. b) Simple criteria to distinguish the two states of chromatin. Heterochromatin is the most repeated fraction of the genome. Constitutive and facultative heterochromatin are included respectively within the HR and MR fractions. Heterochromatin includes highly repeated sequences known as AT rich and gene poor and is thus considered to be a silenced state of chromatin. Euchromatin is mostly included within the SL fraction but can in some cases be part of the MR fraction. Euchromatin is the expressed and gene-rich fraction of the genome; it contains sequences homologous to Expressed Sequence Tags (EST) known to be more GC-rich. It is somehow more difficult to describe the MR fraction which can include both euchromatic and heterochromatic sequences. c) Chromosome schematic of a hypothetical eukaryotic chromosome where heterochromatic sub-domains are depicted. Telomeres (tel), 45S and 5S rDNA (45S and 5S), pericentric repeats (peri), centromere (CEN) and knob are indicated as grey boxes and interrupted by euchromatic segments (in white) which may include dispersed repeated sequences such as transposable elements (vertical bars)

predominant fraction of heterochromatin is found at 45S ribosomal¹ (45S rDNA) loci (that is, the genes encoding the 45S ribosomal RNA precursor molecule) known as the Nucleolus Organizing Regions (NOR), which are at the top of the short arm of chromosomes two and four, near the telomeres. The 45S rRNA genes are organized in large tandem arrays of about 3.5–4.0Mbp and thus constitute the larger part of Arabidopsis heterochromatin (Copenhaver and Pikaard, 1996). Another heterochromatic fraction is derived from centromeres, a chromosome region dedicated to sister chromatid cohesion and to normal chromosome segregation during mitosis and meiosis. In Arabidopsis, centromeres consist of long stretches of about 0.4–1.4Mbp including short tandem repetitive ‘satellite’ DNA, called 180bp repeats and 106B repeats and also known as the Athila retrotransposon (Hosouchi *et al.*, 2002). Flanking regions of centromeres called pericentric regions include other types of repetitive sequences such as transposable elements (Wicker *et al.*, 2007) and 5S ribosomal DNA (5S rDNA) (Tutois *et al.*, 1999). Heterochromatin is also found in telomeres protecting the chromosome ends from deletion upon DNA replication. Telomeres in most plants consist of tandem arrays of simple TTTAGGG repeats (Zellinger and Riha, 2007). Finally knobs, also called interstitial heterochromatin, first observed by Barbara McClintock in maize but present in several other plants, including Arabidopsis, are heterochromatic loci containing transposable elements located on chromosomal arms (Lippman *et al.*, 2004). On the whole, in Arabidopsis, heterochromatin is formed at repetitive sequences clustered mainly at centromere, pericentric regions and NOR (Figure 6.1c).

Transposable elements, rDNA and satellites are common components of heterochromatin in other plant species. They can vary in term of sequences but their properties remain the same. Indeed, most plant centromeres include satellite DNA elements and a survey of these sequences can be found in Wang *et al.* (2009). Heterochromatin content and heterochromatin organization can be very different among eukaryotes and this is especially true for plants where a three-log range of variation in genome size is observed. Over the past decades, although polyploidy and gene duplication have been described as important mechanisms occurring in plants, it turns out that an increase in heterochromatic sequences was the largest contributor to genome size expansion (Gaut and Ross-Ibarra, 2008). Ma and collaborators carried out a comparative analysis of plant centromere organizations including Arabidopsis, *Oriza sativa* and *Zea mays* with a genome size of 150Mb, 370Mb and 2700Mb, respectively (Ma *et al.*, 2007). Centromere sizes among these species do not explain the variation in genome sizes. Alternatively, transposition of mobile DNA on chromosome arms was shown as the main explanation of genome size expansion (Gaut and Ross-Ibarra, 2008). Thus, in many genomes other than Arabidopsis, heterochromatic sequences are not only concentrated to centromeres and rDNA but can also be found in large numbers on chromosome

¹That is, the genes encoding the 45S ribosomal RNA precursor molecule.

arms. This is now well described in cereals such as maize and wheat in which 70–80% of the genome is made of transposable elements (Choulet *et al.*, 2010). Using *Arabidopsis* as a plant model system to study heterochromatin should therefore be carefully considered since it has a very compacted genome in which heterochromatin is mostly centric and pericentric compared to most plant species.

6.1.2 Epigenetic marks

Heterochromatin forms at repeated sequences including rDNA, transposable elements and tandem repeats. These repeated sequences are tightly associated with epigenetic marks which define the heterochromatin state. Main epigenetic modifications such as DNA methylation, histone modifications, non-coding RNA and histone variants are among the major players in the propagation and modulation of the heterochromatic status and will be described in the following section.

6.1.2.1 DNA methylation

In plants, DNA methylation occurs at cytosine residues in three different contexts, CG, CHG and CHH, where H is either A, C or T. DNA methylation depends on a large number of proteins but with central roles for METHYLTRANSFERASE1 (MET1) for CG, SU(VAR)3-9 HOMOLOG4 / KRYPTONITE (SUVH4/KYP) a member of the Su(var)3-9 class of histone methyltransferases, CHROMOMETHYLASE3 (CMT3) for CHG and CHH and DOMAIN REARRANGED METHYLTRANSFERASE2 (DRM2) for CHH. The function of all three methylases is in some way overlapping as DRM2 is a *de novo* DNA methylase for CG, CHG and CHH whereas MET1 and CMT3 respectively maintain methylation at CG and CHG (reviewed in Feng *et al.*, 2010). Many other factors are involved in this process including DECREASE IN DNA METHYLATION1 (DDM1) required for both methylation and histone tail modification at H3K9 (see below). It is worth noting that *met1* and *ddm1* mutants were indeed identified in screens for loss of centromeric-repeat methylation (Vongs *et al.*, 1993; Kankel *et al.*, 2003). Further investigations of those two key players on transposable elements such as the retrotransposon *ÉVADÉ* (EVD), and transposons from the CACTA family (named CACTA because of the occurrence of CATCA sequences at the edges of the elements) or from MUTator-Like Element (MULE) class, which are usually silent, clearly show that they can be reactivated in *met1* and *ddm1* mutants (Mirouze *et al.*, 2009; Tsukahara *et al.*, 2009). On a genome-wide level, 24% of CG, 6.7% of CHG and 1.7% of CHH are methylated. CG, CHG and CHH methylation are enriched in heterochromatic sequences however, with a marked increase for CG methylation in the gene coding sequences (Cokus *et al.*, 2008). Within the gene coding sequences, up to 30% of CG are methylated while less than 1% of the CHG and CHH are methylated (Widman *et al.*, 2009). Conversely heterochromatic sequences are enriched in CHH as a consequence of methylated cytosine deamination in thymine, one of the most frequent point mutations

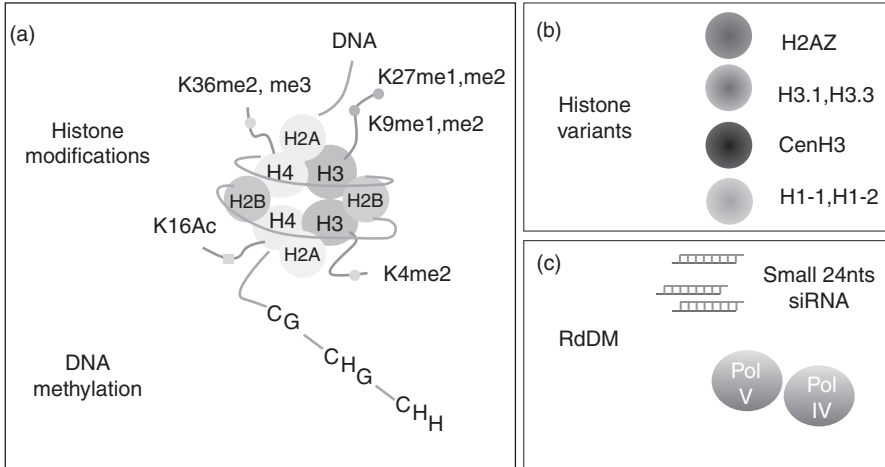


Figure 6.2 Heterochromatin features. a) The nucleosome is the basic structure of chromatin and includes histone octamer H2A, H2B, H3 and H4 as well as DNA. Lysine residues (K) lying within the histone tail at the N-terminal end of histone H3 and H4 can be modified either by methylation (me) or acetylation (Ac). Main repressive marks are indicated in red in Plate 6.2 (H3K9me1, me2 and H3K27me1, me2) whereas activating marks are indicated in green (H3K4me2, H4K36me2, me3 and H4K16Ac). DNA can also be subjected to DNA methylation at CG, CHG and CHH residues. Heterochromatin is enriched in methylated cytosine especially in CHH context, contains histone repressive marks and low levels of histone activating marks. b) Histone variants can substitute for canonical histones. H1-1, H1-2, H2AZ, H3.1, H3.3 and centromeric H3 (homologous to animal CenPA) are discussed in the frame of this review. H2AZ and H3.1 are enriched in actively transcribed regions while H3.3 and centromeric H3 are mostly found within heterochromatic regions. H1 variants are involved in chromosome segregation in plants. c) RNA Directed DNA Methylation (RdDM) occurs at heterochromatic loci and produces small non-coding RNA of 24nts called small interfering RNA (siRNA). RdDM involves plant specific polymerases called Polymerases IV and V (Pol IV and V) and is involved in the establishment of heterochromatin. (For colour details please see colour plate section.)

within eukaryotic genomes (Widman *et al.*, 2009). The overall consequence is that heterochromatic sequences are usually AT-rich sequences which are heavily methylated at CHH (and to a lesser extent at CHG) (Figure 6.1b and 6.2a).

6.1.2.2 Histone code

Core histones H2A, H2B, H3 and H4 are implicated in the formation of the nucleosome core including 147bp of DNA. Histone H1 are linker histones involved in the formation of more condensed chromatin fibres limiting the access for regulatory proteins to nucleosomal components (Happel and Doenecke, 2009). Histones have important implications in gene regulation through modifications at the C-terminal histone tails through methylation, acetylation, phosphorylation or ubiquitination; these

modifications are referred to as the histone code (reviewed in Grant-Downton and Dickinson, 2005).

6.1.2.3 Histone-repressive marks

Heterochromatin-specific histone methylation marks are mainly mono- and dimethyl H3K9 (H3K9me1 and me2) and mono- and dimethyl H3K27 (H3K27me1 and me2) (Figure 6.2a). Specific patterns of histone marks at heterochromatic loci are believed to promote a close chromatin configuration not favourable for gene expression. Basically, two major histone marks are usually followed as heterochromatic marks. These are H3K9me2 implemented by KYP, and H3K27me1 established by histone methyltransferases *ARABIDOPSIS TRITHORAX-RELATED PROTEIN5* and 6 (*ATRX5*, *ATRX6*). H3K9me2 and H3K27me1 are usually found in constitutive heterochromatin but H3K9me2 is dependent upon DNA methylation while H3K27me1 is not, probably reflecting two independent mechanisms occurring at heterochromatic loci (Liu *et al.*, 2010).

6.1.2.4 Histone-activating marks

On the other hand, heterochromatin H3K36me, H3K4me2 and H4K16 acetylation are poorly represented in heterochromatin and mostly found in euchromatic gene-rich regions (Figure 6.2a). H3K36me and H3K4me2 are established respectively by the histone methyltransferases *ABSENT SMALL OR HOMEOTIC DISCS1* (*ASH1*) and *ARABIDOPSIS TRITHORAX1* (*ATX1*) acting mostly at euchromatic loci (reviewed in Liu *et al.*, 2010). Interestingly, histone methyltransferases are unable to methylate target lysine residues that are acetylated, and therefore histone deacetylases (HDAC) are required to allow methylation at those regions (Noma *et al.*, 2001). To date 16 *Arabidopsis* HDACs have been identified and among them HDA6, a homolog of RPD3 in yeast, has a central role in heterochromatin (Pandey *et al.*, 2002). The *Hda6* mutation causes enrichment for euchromatic epigenetic marks such as H3K4 methylation, leading to a decondensed state of chromatin, visible at the cytological level at rDNA loci (Probst *et al.*, 2004). Furthermore, up-regulated loci in *hda6* overlap with those in *met1*, and the *hda6* mutation causes the complete loss of DNA methylation on some HDA6 target loci. These results suggest that HDA6 and MET1 DNA targets are overlapping. Further, this indicates that HDA6 is required together with MET1 to establish and/or maintain heterochromatin (To *et al.*, 2011).

6.1.2.5 Histone variants

Apart from C-terminal core histone modifications, H2A, H2B and H3 histone variants also modulate gene expression (Figure 6.2b). Studies of H3 variants have recently started in plants. Histone H3 variants include H3.1, H3.3 and CENP-A, a centromeric H3 variant, and are encoded by 15 *HISTONE THREE RELATED* (*HTR*) genes (Ingouff and Berger, 2009). These variants could contribute to the epigenetic memory of chromatin states in somatic cells.

Centromeres are epigenetically specified by incorporation of CENH3 (CENP-A in humans; CenH3 is encoded by the *HTR12* gene in Arabidopsis), replacing conventional H3 in centromeric nucleosomes (Henikoff and Dalal, 2005; Ravi and Chan, 2010). Histone variants H2A and H2B were reviewed in Wyrick and Parra (2009) but, to date, very little is known in plants. H2A variant H2A.Z deposition and DNA methylation seem to be mutually exclusive both in actively transcribed genes and in methylated transposons. H2A.Z marks actively transcribed genes and is found at promoter sequences (Zilberman *et al.*, 2008). This may be of particular interest as the yeast H2A histone variant HTZ1 (homologous to plant H2AZ) might tether active promoters to NUP2, a component of the Nuclear Pore Complex (NPC). Recently, it has also been proposed that nucleoplasmic NPC components such as NUP98 and NUP50 can activate expression of internally localized genes, whereas DNA interaction with NUP88 leads to gene silencing (see also Meier this issue; reviewed in Arib and Akhtar, 2011). Future investigations of plant NPC functions will then be of particular interest for gene regulation and heterochromatin.

6.1.2.6 Non-coding RNA

The structural features of heterochromatin led to a simple model where the gene-poor heterochromatin is the condensed state of chromatin and is poorly transcribed while euchromatin is less compacted allowing active transcription of the gene-rich regions. The epigenetic era drastically modified this oversimplified view. Indeed, heterochromatin is transcribed by plant-specific Polymerase II variant enzymes known as Polymerase IV and Polymerase V (Lahmy *et al.*, 2010) and produces large non-coding RNAs (ncRNA) processed through the RNA-directed DNA methylation pathway (RdDM) (Figure 6.2c). This involves ARGONAUTE (AGO) and DICER-LIKE (DCL) families of silencing factors and yields short RNA molecules, 24 nucleotides (24nts) in length, called short-interfering RNA (siRNA) homologous to heterochromatic sequences. The RdDM pathway is a complex network including many proteins and a more detailed description can be found elsewhere (Law and Jacobsen, 2010). SiRNAs are responsible for repression of repeated sequence expression both at the transcriptional and post-transcriptional levels. From an evolutionary point of view, RdDM is believed to be a defence mechanism evolving to protect eukaryotic genomes against mutagenesis by endogenous transposable elements and against exogenous viruses (see discussion in Cavalier-Smith, 2010). As a consequence of the production of 24nts RNA molecules, chromatin is modified both at the DNA level through DNA methylation and at the histone level through post-translational modifications; among them acetylation and methylation are the most studied (see Sections 6.1.2.3 and 6.1.2.4 above; Liu *et al.*, 2010).

6.1.3 Non-histone protein binding

In the early years of the use of Arabidopsis as a plant model species, attempts were made to apply data collected from *Drosophila* and yeast. The best

example related to heterochromatin was the search for Suppressor and Enhancer of variegation (the so called *Su(var)* and *E(var)* genes) involved in the Position Effect Variegation (PEV) in *Drosophila*. In PEV, when a gene is brought close to heterochromatin, this led to its stochastic repression (variegation). In the course of the *Su(var)* and *E(var)* screens, it was demonstrated that PEV was induced by non-histone proteins specific to heterochromatin, which have the property of spreading from heterochromatic to euchromatic loci inducing a silent state of chromatin. Among them, SU(VAR)3-9, a histone methyltransferase establishing methylation at H3K9 and HETEROCHROMATIN PROTEIN 1 (HP1) a highly conserved protein found in yeast, human and plants, were discovered. *Drosophila* HP1 interacts with H3K9me and is a main actor in heterochromatin assembly through interaction with other heterochromatic factors. Once HP1 binds to a given locus, it can induce the spread of heterochromatin in adjacent regions unless a blocking element (called 'boundary elements', see Section 6.2.3.1) blocks heterochromatin propagation (reviewed in Grewal and Jia, 2007). Indeed, plant homologues for SU(VAR) were identified and as an example, SUVH4/KYP (see Section 6.1.2.1) is the functional homologue of SU(VAR)3-9 (Fischer *et al.*, 2006). Arabidopsis HP1, called LIKE HETEROCHROMATIN PROTEIN1 (LHP1) was also discovered but very unexpectedly binds euchromatic loci and recognizes H3K27me₃, a function assumed in *Drosophila* by another class of protein called the Polycomb group proteins. LHP1 should then not be considered as the functional homologue of *Drosophila* HP1 (Gaudin *et al.*, 2001; Turck *et al.*, 2007). Another striking specificity of plant heterochromatin came from the characterization of a duplicated segment of the long arm of chromosome four translocated onto the short arm of chromosome four. This new segment is called hk4S (Lippman *et al.*, 2004). Sequence comparison of the two duplicated segments indicated that hk4S includes the same numbers of genes (33 genes) as the original chromosomal segment but was invaded by 34 retrotransposons and 40 transposons. As expected for heterochromatic sequences, transposable elements at hk4S have high DNA methylation and increased H3K9 methylation while genes are almost unaffected and fully expressed. This was very unexpected and very different from *Drosophila* in which genes would have been repressed by the vicinity of heterochromatic sequences. In plants, spreading of heterochromatin does not seem to occur. Instead, using the *ddm1* and *met1* mutants, Lippman *et al.* (2004) proposed that heterochromatic epigenetic marks are set up by siRNA instead of SU(VAR) in Arabidopsis, in a more accurate and sequence specific manner than the SU(VAR)-mediated process in *Drosophila*.

6.1.4 Heterochromatin is an epigenetic state

An in depth review of heterochromatin can be found in Grewal and Jia (2007) but in summary four main epigenetic features are associated with a repressed chromatin state; DNA methylation (in all three sequence contexts), 24nts siRNA (RdDM), histone methylation at H3K9me₂ and H3K27me₁ and

histone hypoacetylation at H4K16. The exact function of histone variants are as yet largely unknown except for the centromeric H3 variant that will be discussed below. To date there is not a clear understanding of the various steps leading to heterochromatin assembly and this is further complicated by the fact that many processes are interconnected. As an example, DNA methylation, histone modifications and RdDM are tightly linked processes. Indeed if *met1* context is maintained over several generations, RdDM becomes misdirected. The data suggest that CG methylation directs non-CG and H3K9 methylation and is essential for long-term inheritance of epigenetic information (Mathieu *et al.*, 2007).

6.2 Heterochromatin organization

In the previous section, the main properties of heterochromatin were described in terms of sequences and chromatin features. These structural features clearly highlight the fact that heterochromatin distribution is not random but rather clustered at centromeres, pericentromeres, telomeres and knobs although repeated sequences can also be found within euchromatic chromosome arms especially when the genome size increases such as in maize or wheat (Figure 6.1a). However, this linear distribution of heterochromatin along the chromosome does not account for its real organization within the nucleus. Strikingly, the very first investigations of chromosome organization within interphase nuclei in *Arabidopsis* indicated that chromosomes are not arranged in a linear way but rather are distributed within the nucleus with a special location of heterochromatin near the nuclear periphery and around the nucleolus (Fransz *et al.*, 2002). These localizations turned out to reflect specific functional aspects required for proper genome expression. Since then, many nuclear sub-compartments have been described and this nuclear organization of chromatin and sub-compartments are collectively referred to as the 'Nuclear Architecture' (Lanctôt *et al.*, 2007). This section will review our knowledge regarding nuclear architecture focusing on heterochromatin and will give possible mechanisms to explain the positioning of heterochromatin within the nucleus.

6.2.1 Heterochromatin and nuclear architecture

6.2.1.1 Chromosome territories in *Arabidopsis*

Early studies of chromosome spatial organization came from the comparative analysis of active (X_a) and inactive (X_i) X chromosome in humans (Eils *et al.*, 1996). The results suggested that X_a exhibited a larger surface than X_i territories, which should be in a more condensed heterochromatic domain. According to this model, chromosomes would then occupy discrete nuclear domains known as Chromosome Territories (CTs). CTs are an evolutionary conserved structural feature of chromosome organization among plants

and animals and are believed to have functional significance for appropriate genome expression. Chromosome painting was performed on the five *Arabidopsis* chromosomes using 669 Bacterial Artificial Chromosomes (Figure 6.3a). Chromosome Territories were then defined in *Arabidopsis* interphase nuclei in which each CT corresponds to about 25Mbp of DNA, a number very similar to the physical chromosome size ranging from 17 to 25Mbp. Chromosome Territories occur in differentiated and actively dividing cells such as meristematic cells, in various tissues including roots and leaves and in cells of different ploidy levels (Pecinka *et al.*, 2004). Positioning of CTs is random, meaning that there is no specific pattern arrangement in respect to the Nuclear Envelope (NE) or between various CTs except for NOR which are clustered at the nucleolus (Berr and Schubert, 2007; Schubert and Shaw, 2011). The nucleolus was found to be localized in most cases at or near the centre of the nucleus. In addition, the nucleolus represents a significant volume excluding the remaining chromatin at a more peripheral position within the nucleus and thus influences the organization of CTs. In addition, CTs are also constrained by nuclear shape and nuclear size (De Nooijer *et al.*, 2009).

6.2.1.2 Chromocentres and the rosette-loop model of chromatin organization

Each interphase chromosome of *Arabidopsis* consists of a single heterochromatic domain, or chromocentre (CC), from which euchromatic loops of about 0.2–2 Mbp emanate (Van Driel and Fransz, 2004). This has been called the ‘Rosette-loop model’ of heterochromatin (Figure 6.3b). One possible mechanism of loop assembly may rely on pericentric heterochromatin, which includes ancient transposable elements used as a nucleation site to recruit homologous sequences scattered through the chromosome arms (Fransz and De Jong, 2011). The small genome size of *Arabidopsis* provides a unique model species where most of heterochromatin is organized in chromocentres that can be easily followed by light microscopy as dense chromatin foci (Figure 6.3c). Usually, six to ten CCs can be scored in somatic cells because two or more CCs can fuse together (Fransz *et al.*, 2002; Fang and Spector, 2005; Berr and Schubert, 2007). Chromocentres contain epigenetic markers for silent chromatin, such as DNA methylation and repressive histone marks. In contrast, euchromatic loops are enriched in gene-coding regions and contain acetylated histones as well as H3K4me (Fransz *et al.*, 2002). Together, a given CC and associated euchromatic loops form a chromosome territory.

Euchromatin loops can extend from the heterochromatin domains and form either a single rosette structure or multiple rosettes per chromosome, depending on the genome size. In *Arabidopsis*, centromeres are randomly distributed in peripheral positions and near the nucleolus (Fransz *et al.*, 2002). Using centromeric H3 variant CENH3 fused to a fluorescent protein, Fang and Spector (2005) clearly showed that centromeres as part of the CC localize predominantly at the nuclear periphery. Further information on CC assembly was gained by studying heterochromatin in protoplasts. Indeed CCs are

virtually absent from protoplasts except for heterochromatic foci formed at 45S rDNA. Reduction of pericentric heterochromatin is not accompanied by a global loss of histone H3K9me₂, indicating that the H3K9me₂ histone mark is not sufficient on its own to induce compaction of heterochromatin. Following heterochromatin during a longer period of protoplast culture, reformation of CCs could then be observed. The repetitiveness of the sequences explains the kinetics of CC reformation starting with 45S rRNA genes (sub-telomeric NOR), followed by the 180 bp repeats (centromeric repeats) and the 5S rDNA genes (pericentric repeats) (Tessadori *et al.*, 2007). Spatial positioning of Arabidopsis centromeres was similar in several diploid cell types including guard cells, small leaf epidermal pavement cells, root meristematic cells and sepal and petal epidermal pavement cells (Fang and Spector, 2005). However, CC conformation changes have been documented as for instance in endosperm where an Endosperm-Specific Interspersed (ESI) heterochromatin has been described (Baroux *et al.*, 2007). In this tissue, the Rosette model still stands but with a more relaxed conformation. Another situation in which CCs are altered is the use of epigenetic modifier mutations such as *ddm1* and *met1*. In both mutants hypomethylation is observed as well as a reduction of CC size (Soppe *et al.*, 2002). Chromocentres can still be detected in the null *met1-3*

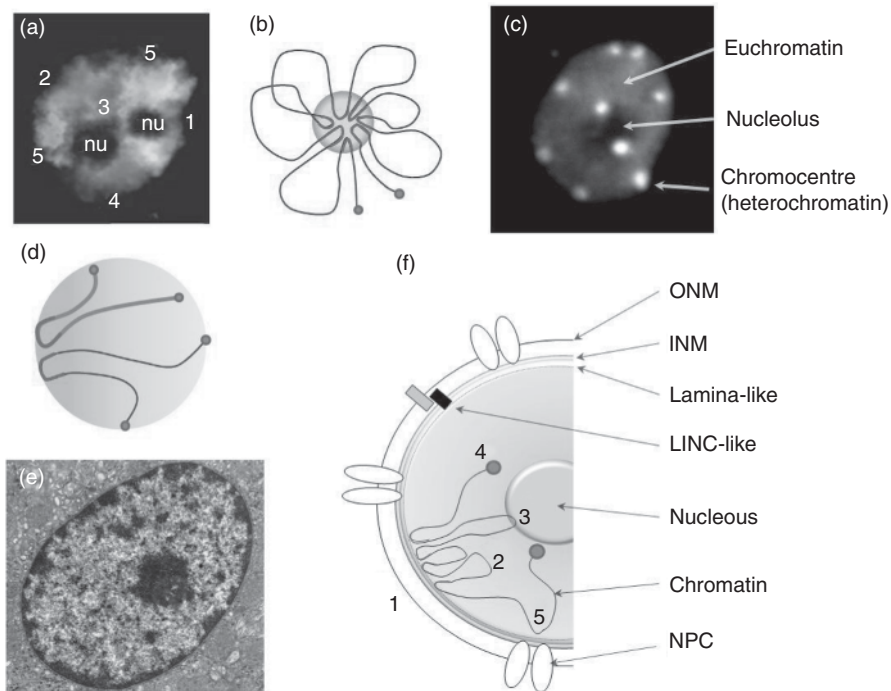


Figure 6.3 (Continued)

mutant, where both H3K9m2 and CG methylation are lacking indicating that those epigenetic marks are not essential for initial heterochromatin assembly (Tariq *et al.*, 2003). Using one specific LacO transgenic line called EL702C and containing three copies of the LacO arrays in two separate loci located on chromosome three, Pecinka *et al.* (2005) found that LacO tandem repeats are often co-localized with chromocentres. This co-localization has also been observed for other tandem arrays such as those of transgenic multicopies

Figure 6.3 Chromatin organization in the nucleus. a) Chromosome Territories (CTs) were identified by FISH experiments using BAC probes against the five chromosomes from *Arabidopsis* with five differential labellings. Chromosome painting was then performed on a tetraploid nucleus including two nucleoli (nu) at interphase stage and revealed discrete CTs corresponding to the five chromosomes (1–5). With kind permission from Springer Science and Business media: Pecinka, A., Schubert, V., Meister, A., Kreth, A.G., Klatte, M., Lysak, M.A., Fuchs, J., 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. b) The rosette-loop chromosome conformation as defined for *Arabidopsis*, from Fransz *et al.* (2002). Heterochromatic loci (in red in Plate 6.3) are grouped together at chromocentres and allow euchromatin to form loops. The chromocentre is indicated as a large yellow circle and corresponds to dense foci observed in 1C. Telomeres are indicated as blue circles. c) *Arabidopsis thaliana* nuclei at interphase from cotyledon stage stained with DAPI (kindly provided by Dr Sylvette Tourmente, <http://www.gred-clermont.fr>). Heterochromatin is more intensely stained due to its large amount of AT-rich sequences and appears as dense foci. Each spot represents a chromocentre although some chromocentres can merge and appear as a single spot. The nucleolus is seen as a non-stained region in the centre of the nucleus. Two chromocentres localize in the vicinity of the nucleolus whereas the six remaining chromocentres are located close to the nuclear periphery. d) The Rab1 chromosome conformation is found in various genomes such as yeast (small genome) or wheat (large genome). Telomeres (blue circle at chromosome extremities in Plate 6.3) and centromeres (indicated in red) are located at opposite sites at the nuclear periphery. e) Electron microscopy image a of nucleus from a parietal stomach cell from *Rattus norvegicus*, showing the nuclear membrane, light euchromatin and dark heterochromatin, some of which is attached to the central nucleolus (Photo credit: image from Dr. Jastrow's electron microscopic atlas on the WWW <http://www.drjastrow.de>). f) Schematic of the chromatin organization in respect to the Nuclear Envelope (NE). The NE is made of an inner nuclear membrane (INM), an outer nuclear membrane (ONM) and contains numerous NPC. A putative plant LINC-like complex is believed to connect the nucleoskeleton and the cytoskeleton. A lamina-like structure has been described in plants although its components remain to be described. Chromatin is represented as a blue line forming rosette-like structures. Heterochromatin (in red) is clustered at the NE (1) whereas euchromatic loops in blue (2) are at more internal locations within the nucleus. Various heterochromatic sequences, such as MARs and TFIIC binding sites may participate in the recruitment of chromatin to the nuclear periphery. Loci encoding transfer RNA or ribosomal RNA (3) and telomeres (4) are brought into, or close to, the nucleolus (large grey circle). Data from other organisms suggest that other loci (5) may be recruited to the NPC where they may be actively transcribed. Panels D, E and F give insights about chromosome organization within the nucleus. (For colour details please see colour plate section.)

of the *HYGROMYCIN PHOSPHOTRANSFERASE (HPT)* locus (Probst *et al.*, 2003). In EL702C, LacO arrays are marked with H3K9me2 but a mutation at *SUVH4/KYP* does not alter its co-localization with the chromocentre (Jovtchev *et al.*, 2011).

Altogether, these studies indicate that H3K9me2 is not responsible for CCs formation. Instead, the repeated nature of heterochromatic sequences, histone variants or siRNA could play an important role in heterochromatin assembly while histone marks and DNA methylation would help to maintain and reinforce heterochromatin compaction. Interestingly, Cavalier-Smith (2010) proposed that heterochromatin first evolved from the characteristic folding of repeated sequences at centromeres. This process would have been subsequently reinforced by the emergence of a specific centromeric histone. siRNA would then be viewed as a mediator of the recognition between heterochromatic sequences and heterochromatin establishing proteins.

6.2.1.3 Chromatin organization in large genome species

Chromocentres are not specific to *Arabidopsis*, although in this species they are easily visualized. Of 67 plants species studied, most display chromocentres in various numbers (Ceccarelli *et al.*, 1998). Organisms with medium-sized genomes (500–3000Mbp), such as tomato contain many heterochromatin domains along their chromosomes. In interphase nuclei, chromosomes form decondensed loops in a rosette-like structure around one or a few number of chromocentres along the chromosome. In mouse, chromocentres are more internally positioned than in *Arabidopsis* (Fedorova and Zink, 2008). In organisms with very large genomes (more than 3000Mpb), such as barley (5000Mbp) and wheat (16000Mbp), in which a large proportion of the genome consists of transposable elements, chromosomes are more condensed. Therefore, the chromocentre model does not really apply in these species (or at least there is a high number of very small chromocentres) and a high number of chromatin loops occur over the complete length of the chromosome (reviewed in Van Driel and Fransz, 2004). In some species or tissues, chromosomes are in 'Rabl' conformation – i.e. telomeres at one side of the nucleus and centromeres at the other side (Rabl, 1885). The Rabl configuration seems to keep the anaphase configuration of chromosomes in a 'frozen' state until the next cell cycle (Figure 6.3d). This organization of the chromosomes clearly supports the existence of links between centromeres and telomeres and the NE. These links will be discussed in the following section. In wheat and its close relatives, the interphase chromosomes adopt a highly regular Rabl configuration with centromeres and telomeres located at opposite poles of the nuclei (Heselop-Harrison *et al.*, 1990). Peter Shaw's group investigated if heterochromatic marks such as methylation and acetylation could affect the Rabl organization of chromosomes. For that purpose they used 5-azacytidine, which reduces DNA methylation, and Trichostatine A, an inhibitor of histone deacetylase (HDAC), and in both cases, the Rabl configuration was not altered, indicating that the links between centromeres, telomeres and the NE

are rather strong (Santos *et al.*, 2002). It is quite difficult to draw a clear picture of why a given species adopts the Rabl conformation, as great variability is observed between species and among tissues or developmental stages. A correlation may exist between chromosome length and organisation where species with large chromosomes adopt the Rabl conformation. Exceptions for this are, however, yeast and *Drosophila*, which have small chromosomes in the Rabl conformation; *Arabidopsis* does not.

6.2.2 Recruitment of heterochromatin at the nuclear periphery

Using electron microscopy, heterochromatin has been defined as a dense structure concentrated close to the NE in most eukaryotic nuclei including plant and animals as illustrated in Figure 6.3e. In *Arabidopsis*, chromocentres clearly co-localize with the NE whereas some are found at the vicinity of the nucleolus (Figure 6.3b). Although preferential localization to the nucleolus has been ascribed to the presence of the NOR at the tips of chromosome two and four (reviewed in Fransz and De Jong, 2011), little is known about the preferential localization of the remaining chromocentres at the NE. Links between heterochromatin and NE should then exist and the NE has a special place within this review as it is believed to participate in chromatin anchorage in plants. A more exhaustive description of NE composition and structure is given in Chapter 2 and the NPC is described in Chapter 3. This section will focus on chromatin-NE interactions. Over the past decades, many interactions involving NE and DNA/chromatin have been described, especially in animals, but with similarities in plants. These data may provide insights into possible mechanisms for maintaining heterochromatin close to the NE in plants.

6.2.2.1 The central role of lamins in animals

From a cytological point of view, heterochromatin is in the immediate vicinity of the internal face of the NE (the Inner Nuclear Membrane, INM) and in contact with the lamina, a thin meshwork made of lamins A, B and C (Figure 6.3f). Lamins belong to the large coiled-coil family of proteins and form a scaffold, which connects the various NPCs and lends mechanical strength to the nucleus. In humans, the premature aging disease Hutchinson–Gilford progeria syndrome (HGPS) is caused by a mutant lamin A (LA Δ 50). In LA Δ 50, nuclei are abnormally shaped and display a loss of heterochromatin. In humans, H3K27me₃ and H3K9me₃ identify facultative heterochromatin, whereas H4K20me₃ marks constitutive heterochromatin. Consistently with the loss of heterochromatin, constitutive and facultative heterochromatic marks are altered in LA Δ 50 (Shumaker *et al.*, 2006). *In vitro* lamin B has some affinity to H2A but not to the other core histones (see Mattout *et al.*, 2007 and references herein) and interactions between lamins and chromatin may involve other protein factors (see below).

6.2.2.2 The inner nuclear membrane and heterochromatin

At least 80 unique proteins have been found localized to the NE in animal cells (Tzur *et al.*, 2006), but few of them have been well characterized. These include KLARSICHT/ANC-1/SYNE HOMOLOGY (KASH), LAMIN B RECEPTOR (LBR), LEM proteins (LAMINA-ASSOCIATED POLYPEPTIDE-EMERIN-MAN1), SAD-1/UNC-84 (SUN) and NESPRIN (see Chapter 2 in this volume). LBR is an integral NE protein interacting both with lamins and heterochromatin. It has a complex function and structure that have been recently reviewed in Olins *et al.* (2010). *In vitro* it binds directly to DNA within the linker region between nucleosomes and interacts with a large number of partners including HP1 and Methyl Binding Protein 2 (MeCP2) (Olins *et al.*, 2010). HP1 has a central role in the assembly of heterochromatin in other organisms such as yeast and *Drosophila* where it recognizes histone H3K9 methylation while MeCP2 recognizes methylated cytosine at the DNA level. Strikingly, the LBR loss of function led to HP1 α mislocalization. Whether this is a direct or indirect effect remains to be determined. As previously stated (see Section 6.2.1.3), chromosome organization in mouse chromocentre-like structures recalls the rosette-loop model adopted by the *Arabidopsis* chromosomes. Indeed, the LBR knock-out mouse displays a reduced number of chromocentre-like structures, which are more diffused suggesting a more relaxed state of heterochromatin (Cohen *et al.*, 2008).

Another interesting NE component was shown to directly interact with centromeric heterochromatin in fission yeast *Schizosaccharomyces pombe*. In this species, centromeres are connected to the primary microtubule organizing centre (MTOC; or spindle pole body (SPB), equivalent to the animal centrosome) through interactions between a SUN domain protein called SPINDLE POLE BODY-ASSOCIATED PROTEIN1 (SAD-1) and two KASH domain proteins KARYOGAMY MEIOTIC SEGREGATION PROTEIN1 and 2 (KMS1 and KMS2). Connections are stabilized by the INM protein ISOMALTASE 1 (IMA1) (reviewed in Mekhail and Moazed, 2010). SUN and KASH proteins are part of the Linker of Nucleoskeleton and Cytoskeleton (LINC) complex connecting the cytoskeleton and the nucleoskeleton (see Chapter 2; see also Figure 6.3f in the current chapter). Recently, the first KASH protein has been discovered in plants and has been shown to interact with SUN domain proteins (Zhou *et al.*, 2012). Further investigations will be needed to establish if SUN domain proteins or other LINC-like complexes are used in *Arabidopsis* to link the NE and the chromocentres, offering one possible mechanism of chromocentre positioning in *Arabidopsis*.

6.2.2.3 Heterochromatin positioning in plants

In plants, no functional lamina homologue has been described so far, although heterochromatin still localizes peripherally. Electron microscopy performed on tobacco BY-2 cells identified a filamentous structure closely attached to the INM and linked to NPCs. The organization and dimensions of these filaments are very similar to those observed for the lamina in *Xenopus laevis* oocytes used as a control in the experiments (Fiserova *et al.*, 2009). These data suggest

the existence of a lamina-like structure in plants but are not a definite proof of its existence. As lamins belong to the coiled-coil family, several researchers then investigated this large gene family in plants. In tomato, the Meier group identified MATRIX ATTACHMENT REGION (MAR) BINDING FILAMENT-LIKE PROTEIN1 (MFP1, Gindullis and Meier, 1999a), MFP1 ASSOCIATED FACTOR1 (MAF1, Gindullis *et al.* 1999b) a serine threonine-rich protein interacting with MFP1 and the FILAMENT-LIKE PLANT proteins (FPP family, Gindullis *et al.*, 2002) also interacting with MAF1. MAF1 was subsequently shown to be the first WPP-domain protein conserved in Arabidopsis. However, WPP-domain proteins were recently shown to be part of the plant NPC and not directly connected with heterochromatin (Meier *et al.*, 2010). A similar approach was developed in carrot, in which a coiled-coil protein called NUCLEAR MATRIX CONTTTUENT PROTEIN1 (NMCP1) was identified as a plant-specific insoluble nuclear protein localized at the nuclear periphery. Using NMCP1 as a starting point for a reverse genetics screen, Dittmer *et al.* (2007) identified a new family of proteins called LITTLE NUCLEI 1-4 also known as CROWDED NUCLEI1-4 (LINC1-4, CRWN1-4) in Arabidopsis. Disruption of either the *LINC1* or *LINC2* gene induces a dwarf phenotype and causes a reduction in nuclear size associated with an increased DNA compaction. Moreover, combining the *linc1* and *linc2* mutations reduced by two-fold the number of chromocentres (8.6 compared to 4.6 chromocentres per nucleus in wild type and double mutant respectively). This suggests that barriers against association between distinct chromocentres are lowered, resulting in the fusion of some of the chromocentres. Altogether, the *linc1 linc2* double mutant displays a reduced nuclear size, an increased DNA compaction and a reduction of the number of chromocentres. One possible explanation could be that the nuclear shape alters chromatin organization or *vice versa*. Although the function and structure of the LINC proteins may recall those of animal lamins, further work on LINC proteins is needed and will determine whether they influence NE structure or modify links between chromatin and NE.

6.2.3 Higher order of chromatin organization

Spatial positioning of heterochromatin in rosette-loops at the chromocentre suggests that heterochromatin can bring distant loci closer together. For Arabidopsis, one possible mechanism has already been proposed in Section 6.2.1.2 in which transposable elements on chromosome arms are recruited to the chromocentre (Fransz and De Jong, 2011). In other species, dedicated sequences and proteins that can fold chromatin into loops were identified with similar properties. In *Drosophila*, boundary elements like CCCTC-BINDING FACTOR (CTCF), SUPPRESSOR OF HAIRY-WING (Su(Hw)), ZESTE WHITE 5 (ZW5), BOUNDARY ELEMENT ASSOCIATED FACTOR (BEAF) or CENTROSOMAL PROTEIN 190 (CP190) are involved in chromatin organization (Bushey *et al.*, 2008) and some of them have been implicated in chromatin loop formation (Blanton *et al.*, 2003). Some of these proteins were

hypothesized to interact with the nuclear lamina, recruiting the boundary elements in a way very similar to that described for heterochromatic sequences of the rosette-loop model (Gerasimova *et al.*, 2000). Boundary elements and their binding partners are thus among the possible mechanisms responsible for chromatin loop formation and defining chromosome territories in eukaryotes.

6.2.3.1 Boundary elements

6.2.3.1.1 *Drosophila* and animal CTCF In order to recover an enriched fraction of NE-associated chromatin, lamin was used in human and drosophila cell cultures as a fusion with Dam methylase from *Escherichia coli*, the so-called 'DamID' method (Van Steensel *et al.*, 2001). Dam methylase targets adenines while the lamin counterpart tethers the fusion protein to the NE and binds specific DNA sequences. This led to the discovery of Lamin-Associated Domains (LADs) of about 0.5 to 1Mpb each (Pickersgill *et al.*, 2006; Guelen *et al.*, 2008). In *Drosophila* Kc cells, these domains contain essentially silent genes characterized by a low level of H3K4 methylation and low H4K16 acetylation (see Figure 6.2a). Careful analyses revealed that CTCF-binding sites are at the edges of LADs (Pickersgill *et al.*, 2006). CTCF is a zinc finger protein binding CCCTC motifs; it is well known as both an insulator (enhancer-blocking property) and a boundary element (blocking repressive effects mediated by heterochromatin) (reviewed in Gaszner and Felsenfeld, 2006). To date no CTCF homologues have been identified in plants

6.2.3.1.2 *Transcription factor III C* How do organisms that lack CTCF homologues accomplish the same goal? In yeast, the transcription factor III C (TFIIC) could have a similar function. TFIIC is an associated factor of RNA Polymerase III (Pol III), which has been linked to insulator activity. TFIIC binds B-Box elements found in Pol III promoters such as in tDNA, rDNA, U6snDNA or SINE transposable elements all transcribed at the nucleolus. TFIIC recognizes the 274 tRNA gene clusters distributed throughout the 16 chromosomes and could contribute significantly to organizing the yeast genome by tethering tRNA gene clusters to the nucleolus. This can be viewed as an alternative mechanism to establish or maintain loops, which bring loci from distinct chromosomal positions closer together. As an example, in fission yeast, transition from CENP-A centric heterochromatin to outside repeats coincides with the presence of two to four tRNA genes, which may define boundaries of centromeres (reviewed in White and Allshire, 2008).

Genome-wide studies identified a second type of TFIIC binding site not associated with RNA Pol III binding. While RNA Pol III-associated TFIIC loci are recruited at the nucleolus, such sites are most often anchored to the NE and defined as 'Chromosome Organizing Clamp' (COC) (Noma *et al.*, 2006). The recent discovery of hundreds (possibly even thousands) of COC sites in the human genome points towards an important, conserved function for these sites in organizing eukaryotic genomes (Moqtaderi *et al.*, 2010; Oler *et al.*, 2010).

6.2.3.2 Condensin and cohesin

Condensins and cohesins, two chromosome scaffold proteins, are members of the Structural Maintenance of Chromosomes (SMC) family of proteins. They act respectively on chromosome compaction prior to and during mitosis and cohesion between the sister chromatids upon replication during the S phase until mitosis (reviewed in Wood *et al.*, 2010). Arabidopsis cohesins (SMC1 and 3) and condensins (SMC2 and 4) have been identified and mutants in SMC1, 2 and 3 display a similar 'Titan' (TTN) phenotype: giant endosperm nuclei and arrested embryos with a few small cells (Liu *et al.*, 2002). One of the functional roles of TTN is to provide normal microtubule function during seed development, which may explain the enlargement of the nuclei in this tissue. Strikingly, DMS3, another Arabidopsis SMC-like protein, was shown to be involved in the RdDM connecting proteins with nuclear scaffold function and epigenetic regulation (Kanno *et al.*, 2008).

In animals, SMC complexes are bound to chromatin all through the cell cycle and recent investigations have highlighted implications of this binding in some aspects of genome organization. First, cohesin was shown to co-localize with CTCF binding sites. It is even hypothesized that the boundary function first described for CTCF may be achieved by cohesin. Cohesin may be recruited at CTCF binding sites and subsequently bring together unlinked loci to form DNA loops. The current hypothesis would then be that CTCF defines binding sites for cohesin, which in turn induces DNA topology responsible for insulator and/or boundary effects (Wendt *et al.*, 2008). Second, genome-wide studies identified condensin-binding sites at tRNA loci. A mutation in the condensin sub-units induces a loss of preferential localization of tRNA loci to the nucleolus. As for cohesin, it is suggested that condensin facilitates tRNA clustering to the nucleolus by participating in long range interactions between distant chromosome sites (D'Ambrosio *et al.*, 2008).

6.2.3.3 Matrix Attachment Regions

Matrix Attachment Regions (MARs) are A-T rich repeated DNA sequences found in animals and plants. At least 12 proteins were described as binding at MARs in various organisms, including lamins, histone H1, topoisomerase II, as well as plant-specific factors already discussed such as MFP1 and MAF1, which are components of the NE (Wang *et al.*, 2010). In *Mus musculus*, another MAR binding factor called SPECIAL AT-RICH BINDING PROTEIN1 (SATB1) is responsible for repression of numerous genes mediated by deacetylation at histone H3K9 meaning that MARs can also be defined at the epigenetics level (Cai *et al.*, 2003). MARs are attached to the nuclear scaffold and are believed to organize metaphase chromosomes into rosette-like structures forming loops. It is worth noting that transposable elements can be recovered, to a certain extent, as MARs (Tikhonov *et al.*, 2001) according to their possible function in the rosette-loop model in Arabidopsis. The attachment sites of the loops are hypersensitive to DNaseI treatment and co-localize with Topoisomerase II binding sites. Using these two criteria, MARs were identified in plants in

which they form loops of various sizes ranging from 25 kb in *Arabidopsis* to 45 kb in maize (Paul and Ferl, 1998). *In silico* analyses predicted smaller loops around 6–7 kpb as 21705 MAR sequences were detected across the *Arabidopsis* genome by Rudd *et al.* (2004). Among them, 10% lie within genes where they may influence gene expression and confer tissue, organ, and developmental specificity of gene expression in *Arabidopsis* (Tetko *et al.*, 2006). Finally, in vertebrates, the β -globin insulator element that is binding CTCF can also be recovered as a MAR (Yusufzai and Felsenfeld, 2004). The authors clearly ruled out the fact that all the sequences isolated as MARs are also insulator elements (there can be as many as 30 000 to 80 000 MARs in the human genome). Alternatively, CTCF may have some yet unknown function in MARs. The exact significance of MARs is still elusive and their effective relationships with the rosette-loop model in *Arabidopsis* remains to be defined.

6.2.3.4 Future prospects in plants

In summary, interactions of heterochromatin to form chromatin loops may rely on putative structural elements such as transposable elements, boundary elements, MARs, or on protein complexes such as SMC or TFIIIC complexes or by directly interacting with the NE or through a lamina-like structure (Figure 6.3f). In plants, no boundary or insulating elements similar to CTCF or other *Drosophila* insulators have been identified. Boundary elements may not be needed and alternatively, as proposed by Lippman *et al.* (2004), siRNA-dependent mechanism may be sufficient to fulfil the function of a boundary element by defining well delimited chromatin domains.

6.3 Functional significance of heterochromatin positioning

The structure of heterochromatin led us to discover that it is composed of repeated sequences and constitutes a specific chromatin state distinct from euchromatin. The study of its organization has revealed that heterochromatin is grouped in specific regions of the chromosome and that this organization may participate in the positioning of heterochromatin at the periphery of the nucleus and euchromatin at a more internal part of the nucleus. If this positioning is a biological reality, it is probably relevant for some functional properties of the cell. In the following section, three main roles in cell functions will be reviewed: chromosome segregation, transcriptional regulation and genome protection against instability. Whether these functions are dependent upon the peripheral positioning of heterochromatin will be discussed.

6.3.1 Centric heterochromatin directs chromosome segregation

Clearly, centromere positioning at the nuclear periphery is an essential process to achieve chromosome segregation. Thus mechanisms dragging

chromosomes to each side of the cell can yield clues about one of the possible mechanisms for positioning chromatin at the nuclear periphery. At the sequence level, centromere identity relies on particular chromatin specificities made of dedicated repeated sequences (satellite DNA in plants) and involving small non-coding RNA and specific histone variants such as CenH3/CENP-A and H3.1.

During mitosis and meiosis, centromeric heterochromatin is attached to spindle microtubules through the kinetochore, a proteinaceous structure connecting the centromere and the spindle. CENP-A together with CENP-C are required to build up a fully functioning kinetochore and thus have a central role in the centromeric function in chromosome segregation (Verdaasdonk and Bloom, 2011). In *Arabidopsis*, the 180bp satellite is believed to define a key centromeric feature of all centromeres. Indeed 180bp arrays bind the centromeric histone variant CenH3 but in extended chromatin fibres studied by FISH, only 10–12% of the repeats bind CenH3. These were compacted chromatin regions forming knobs and were hypothesized to be sites of kinetochore formation (Shibata and Murata, 2004). What makes some of the 180bp competent to bind CenH3 and to form kinetochores remains to be elucidated. CenH3 mutants have been recently obtained and this will provide the opportunity to investigate the functions of the centromeric histone variant in chromocentre positioning (Lermontova *et al.*, 2011).

Once kinetochores are formed, they are connected to spindle microtubules and allow chromosome segregation. It is worth noting that, in plants, spindle attachment does not rely on specialized organelles (centrosomes in animals). Instead, plants are believed to follow a specific mechanism of spindle formation known as ‘spindle self-organization’ relying on the Ran pathway (reviewed in Zhang and Dawe, 2011). Surprisingly, in this model, Histone H1 acts as a microtubule organizing factor at the nuclear periphery and indeed Histone H1 was shown in tobacco BY2 cells to be localized at the nuclear periphery (Nakayama *et al.*, 2008). The identified histone H1 in this experiment was shown to be similar to one of the two main histone variants from tobacco BY2 cells. The authors observed that histone H1 and DNA do not co-localize suggesting that the recognized histone H1 variant may, in this case, have a distinct function from the classical linker histone association.

In *Arabidopsis*, H1 variants were characterized and shown to be encoded by three genes. *H1-1* and *H1-2* share extensive homology with each other whereas *H1-3* is a more divergent variant related to a drought-inducible class of gene (Wierzbicki and Jerzmanowski, 2005). RNAi plants with ~90% decrease in the overall level of histone H1 display developmental phenotypes, with stochastic alterations of DNA methylation at various loci including 180bp, 5S rDNA, transposable element and FLOWERING WAGENINGEN (FWA). It is currently difficult to reconcile the two reported functions of histone H1, namely linker histone involved in chromatin compaction (reviewed in Jerzmanowski, 2007) and its microtubule organizing function (Wierzbicki and Jerzmanowski, 2005). This may be solved in the future by carefully

looking at each H1 variant to address its specific functions and localization within the nucleus.

In most plant species, centromeres are close to the NE whether or not they display the Rabl configuration. This close proximity to the NE favours rapid interaction with microtubules during NE breakdown. Further investigations are needed to define clearly the function of histone variants and ncRNA or unknown factors in centromere positioning. This may be of importance for deciphering one of the possible mechanisms to achieve heterochromatin positioning.

6.3.2 Spatial positioning of heterochromatin affects transcriptional activity

As previously discussed, chromosomes are spatially organized in interphase nuclei and heterochromatin tends to be situated next to the NE and the nucleolus (see Figure 6.3). As heterochromatin represents a repressive state of chromatin, it has long been suggested that transcription may be repressed at the NE and this has been well described for telomeric silencing in yeast (Akhtar and Gasser, 2007). Yeast and mammals were largely used to investigate the effect of heterochromatin on gene expression but also the effect of gene tethering to the nuclear periphery or to the nucleolus (Towbin *et al.*, 2009; Sáez-Vásquez and Gadal, 2010). However this turned out to be a difficult task, and the effect on transcription of the perinuclear positioning of chromatin is indeed highly dependent upon the chosen locus, how it is recruited to the nuclear periphery and how it is positioned relative to various NE components such as the NPC. What about plants? One example in plants came from the study of *FWA* gene (Soppe *et al.*, 2000). In its wild-type configuration, *FWA* is methylated at the DNA level whereas mutants *fwa-1* and *fwa-2* display a complete loss of DNA methylation associated with a late-flowering phenotype. *FWA* contains in its promoter region some sequence repeats that were mapped as the key determinant of this epigenetic regulation. It is interesting to note that *FWA* is among the 33 genes of the *hk4S* translocation (see Section 6.1.3). As described above, most genes included within *hk4S* are not altered by the spreading of repressive marks from the nearby transposable elements as only very slight differences can be recorded in their DNA methylation profiles. However, using the *ddm1* mutant background, Lippman *et al.* (2004) identified a derepression of *FWA* at the transcriptional level, indicating that genes can be modulated by DNA repeats such as transposable elements but only when they are located close to their transcription units. Is positioning of *FWA* responsible for this derepression? The position (inside or outside a given chromosome territory) of the *FWA* gene was then determined when *FWA* was in the active (*ddm1* background) and in the silent state (*wild type* background). Unfortunately, no significant difference in the nuclear position of the *FWA* locus could be identified (Pecinka *et al.*, 2004). Of course this is only one example and, as stated before, the effect of spatial positioning may

not apply to every gene or may be dependent upon chromosomal position or sequence environment.

In an attempt to get a more general tool to investigate the function of spatial positioning in gene activity, Eric Lam's group engineered a set of 277 transgenic lines containing a lac operator (LacO) and expressing Luciferase under the CaMV 35S promoter (Rosin *et al.*, 2008). Each locus has been characterized for its chromosomal position and its level of gene expression through the LUC activity. LacO arrays can then be targeted using fluorescent protein fused to the Lac repressor (LacI) in order to track chromosome positioning (a given LacO transgene) in living plants. A similar approach has also been developed by the Matzke group using the tet operator (Matzke *et al.*, 2005). Using *ddm1*, *met1* and 5-azacytidine, they showed that changing the epigenetic status can, for some transgenes, decrease the Luciferase activity and modify the nuclear position of the LacO transgene within the nucleus (Rosin *et al.*, 2008). However, this is not a general rule and the effect differs depending on the chromosomal position of the transgene array (reviewed in Lam *et al.*, 2009).

Gene regulation by spatial positioning is a new field of investigation in plants. Tools to study transgene tethering to sub-nuclear localizations are becoming available and future investigations will be needed to establish whether the nuclear periphery induces a repressive or activation effect in respect to gene expression.

6.3.3 Heterochromatin positioning protects against genome instability

In most organisms, pericentric heterochromatin contains inactive transposable elements that are used to silence the homologous euchromatic copies. Transposable elements are then mainly in a repressed state, controlled by the RdDM pathway in plants. One clear consequence of this genomic organization is that pericentric heterochromatic repeats protect the genome against transposable element mobilization, which would otherwise lead to genome mutagenesis. The repressed state can be released in some epigenetic mutants such as *ddm1* or *met1*, in which transposable element mobilization can then be observed (Mirouze *et al.*, 2009; Tsukahara *et al.*, 2009). In *ddm1*, there is a clear correlation between heterochromatin decondensation illustrated by a reduced number of chromocentres and derepression of heterochromatic sequences.

One striking feature of heterochromatin not yet investigated in plants is that of rDNA as highlighted by recent yeast studies. This is an essential gene encoding ribosomal RNA but it also belongs to heterochromatic sequences due to its tandem repeat organization (see Section 6.1 and Figure 6.1). rDNA is found in all organisms from prokaryotes to eukaryotes where it is organized in tandem arrays, meaning that a gene amplification system maintains rDNA cluster(s) in a tandemly repeated organization. Each species displays

a given number of copies: seven in *Escherichia coli*, 150 in *Saccharomyces cerevisiae*, 350 in *Homo sapiens*, at least 1000 in *Arabidopsis* and up to 12 000 in *Zea mays*. rDNA arrays are an evolutionary conserved process to produce a large amount of rRNA without increasing the transcription rate of a given gene but surprisingly not all the copies within an array are transcribed. Why are untranscribed units kept within the rDNA array? Unexpectedly, deletion of untranscribed rDNA copies induces an increased sensitivity of DNA to mutagenic agents. Silent extra rDNA copies were shown to facilitate condensin association and sister-chromatid cohesion, thereby facilitating recombination repair between sister-chromatids. Losing the extra copy thus reduces the capacity of DNA repair using the sister-chromatid. In *Arabidopsis*, chromosome pairing between sister-chromatids occurs at transgene repeats such as LacO or HPT. Homologous pairing is correlated with DNA hypermethylation at the array (Watanabe *et al.*, 2005). Yeast genetics yielded one more role of condensin, which not only functions in DNA compaction but is also required for the attachment of rDNA arrays from the sister-chromatids to each other with important implication in genome stability. This remains to be studied in plants.

6.4 Perspectives

The definition of heterochromatin relies on various criteria. In this review, three main criteria were used: structure, organization and biological functions (Table 6.1). Despite this, it is difficult to establish a general definition of heterochromatin that would suit all situations encountered. Some scientists will use the repeated nature of DNA sequences, its position on the chromosome, a combination of epigenetic marks or a sub-nuclear localization of chromatin. However, none of them are sufficient by themselves to define the heterochromatin state of chromatin. As an example, repeated gene families such as the *NUCLEOTIDE-BINDING SITE-LEUCINE-RICH REPEAT* (NBS-LRR) disease resistance genes (150 genes in *Arabidopsis*, 400 in rice – McHale *et al.*, 2006) are organized in clusters but cannot be considered as heterochromatic loci; methylation at H3K9 can also be found at euchromatic loci and cannot be strictly assigned to heterochromatin; and recruitment at the nuclear periphery does not always lead to repression as actively transcribed genes may be recruited to NPCs. When considering well established heterochromatic sequences such as the pericentric 5S rDNA repeats at chromosome five, some repeats are clearly expressed whereas others remain repressed. Active and repressed 5S genes are associated respectively with active and repressive epigenetic marks although all these genes are within the same repeated array (Mathieu *et al.*, 2003). Finally, it remains to be elucidated how to define heterochromatin in species with large genomes, such as wheat, in which a large number of transposable elements are scattered along the

Table 6.1 Criteria used in the definition of heterochromatin

Criteria	Heterochromatin Features	Heterochromatin Characteristics	Short Description	References
Structure	DNA sequences	Gene poor Contains satellites Contains transposable elements	Heterochromatin is made of repeated sequences 180bp repeats in Arabidopsis Mostly pericentric in Arabidopsis	Cokus <i>et al.</i> (2008) Hosuchi <i>et al.</i> (2002) Arabidopsis Genome Initiative (2000)
		May form knobs Includes 5S and 45S rDNA	hk4S knob in Arabidopsis Subtelomeric 45S and pericentric 5SrDNA	Lippman <i>et al.</i> (2004) Copenhaver & Pikaard (1996); Tutois <i>et al.</i> (1999)
	Epigenetic marks	High DNA methylation	CG, CHG and CHH methylation at cytosine residues Enriched H3K9me2, H3K27me1	Cokus <i>et al.</i> (2008) Liu <i>et al.</i> (2010)
		Contains repressive histones marks Depleted in activating histone marks	Decreased H3K4me2, H4K16Ac	Probst <i>et al.</i> (2004)
		Contains specific Histone variants Is targeted by non-coding RNA	H3.1, CenH3 Centromeric variant (CenPA homolog) Heterochromatin targeting by RNA Directed DNA Methylation (RdDM)	Ingouff et Berger (2009) Law and Jacobsen (2010)
	Non-histone protein binding	Binds specific Su(var) proteins	Arabidopsis LHP1 has functions distinct from animal HP1	Gaudin <i>et al.</i> (2001)

(continued)

Table 6.1 (Continued)

Criteria	Heterochromatin Features	Heterochromatin Characteristics	Short Description	References
Organization	Chromosome distribution Nuclear architecture	Located at centromere, telomere, pericentromeres, knob Involved in Chromosome Territories Forms chromocentre vs Rabl conformations Enriched at nuclear periphery Under represented at NPC	See DNA sequences Arabidopsis Chromosomes are organized in CTs Arabidopsis Chromosomes form chromocentres Repressive context for transcription Activating context for transcription in yeast To be investigated in Arabidopsis	Arabidopsis Genome Initiative (2000) Pecinka <i>et al.</i> (2005) De Jong and Franz (2011) Fang and Spector (2005); Andrey <i>et al.</i> (2010) Arib and Akhtar (2011) –
Functions	Higher order of chromatin Chromosome segregation Transcription regulation Protection against genome instability	May contain boundary elements, MARS Forms centromere Is a silenced state of chromatin Repressed transposition and recombination	Satellite, histone variants and epigenetic state are involved Epigenetic repressive marks, chromatin compaction Transposable elements, rDNA (in yeast)	Zhang and Dawe (2011) Lam <i>et al.</i> (2009) Mirouze <i>et al.</i> (2002); Watanabe (2005)

chromosome arms. It is therefore important to direct future research towards better defining heterochromatin. For instance, in *Drosophila*, sub-classes in relation to the various criteria set out above have been established (Filion *et al.*, 2010).

While heterochromatin components are well documented, the chronological events leading to its assembly are still largely unknown. Clearly, several studies suggest that histone methylation at H3K9 is not needed to build heterochromatin (Tariq *et al.*, 2003; Tessadori *et al.*, 2007; Jovtchev *et al.*, 2011). Cavalier-Smith (2010) proposed that centric heterochromatin evolved first, the repeated nature being the primary determinant of heterochromatin subsequently stabilized by centrosomal histone variants CENP-A. According to this hypothesis, plasmid partitioning in prokaryotes involves three simple components: a repeated DNA sequence (centromeric repeats in eukaryotes), a centromere binding protein (histone variants in eukaryotes), and an associated motor protein (microtubules in eukaryotes) to separate the segregating plasmids (Wilson and Dawson, 2011). In plants, it will be very relevant to study the role of the various histone variants to determine whether they are needed in the initial steps of heterochromatin assembly.

Nuclear architecture is a recent field of investigation in plants and is hypothesized to participate in the regulation of genome expression. Heterochromatin may be an important component in this mechanism by anchoring chromatin loops, a mechanism that nicely explains the formation of chromocentres in *Arabidopsis* (Fransz and De Jong, 2011). Chromocentre positioning falls into two main classes as some chromocentres cluster at the nucleolus while others are in close proximity to the NE. Nucleolus association can be explained by the fact that 45S rDNA repeats have to be actively transcribed within the nucleolus and, as a more general rule, genes transcribed by RNA Pol III and Pol I may be recruited to the nucleolus (Sáez-Vásquez and Gadal, 2010). However, preferential association of chromocentres at the nuclear periphery remains elusive. This may be reminiscent of chromosome segregation or other functions discussed in Section 6.3. Further, mechanisms responsible for chromocentre positioning still remain to be discovered in plants. Collectively, authors suggest that the nucleolus, because of its central position, excludes the chromocentres from internal localization except those bearing NOR. Chromosome territories were shown to be randomly distributed but form distinct areas within the nucleus suggesting a mutual exclusion between different territories (Pecinka *et al.*, 2004; Berr and Schubert, 2007; De Nooijer *et al.*, 2009; Andrey *et al.*, 2010). Data coming from other species suggest specific interactions between NE components and heterochromatin, providing an alternative hypothesis that could explain heterochromatin positioning. Many issues are yet unsolved: lamin and LBR are not yet described in plants and KASH proteins have only recently been discovered. The possible implication of SUN and LINC proteins in NE-heterochromatin interactions also remain to be established.

In most species heterochromatin is the main form of chromatin. Its functional significance is constantly growing as our understanding of the cell process is increasing. Research into its assembly, maintenance and positioning will give us a better view of the processes involved in the regulation of genome expression.

Acknowledgements

The authors thank Aline Probst and David Evans for helpful discussions and helpful comments on the manuscript.

References

- Akhtar, A. and Gasser, S.M. (2007) The nuclear envelope and transcriptional control. *Nature Reviews Genetics* **8**, 507–517.
- Andrey, P., Kiêu, K., Kress, C. *et al.* (2010) Statistical analysis of 3D images detects regular spatial distributions of centromeres and chromocenters in animal and plant nuclei. *PLOS Computational Biology* **6**, e1000853.
- The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
- Arib, G. and Akhtar, A. (2011) Multiple facets of nuclear periphery in gene expression control. *Current Opinion in Cell Biology* **23**, 346–353.
- Baroux, C., Pecinka, A., Fuchs, J. *et al.* (2007) The triploid endosperm genome of *Arabidopsis* adopts a peculiar, parental-dosage-dependent chromatin organization. *Plant Cell* **19**, 1782–1794.
- Berr, A. and Schubert, I. (2007) Interphase chromosome arrangement in *Arabidopsis thaliana* is similar in differentiated and meristematic tissues and shows a transient mirror symmetry after nuclear division. *Genetics* **176**, 853–863.
- Blanton, J., Gaszner, M. and Schedl, P. (2003) Protein:protein interactions and the pairing of boundary elements in vivo. *Genes and Development* **17**, 664–675.
- Britten, R.J. and Kohne, D.E. (1968) Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science* **161**, 529–540.
- Brown, S.W. (1966) Heterochromatin. *Science* **151**, 417–425.
- Bushey, A.M., Dorman, E.R. and Corces, V.G. (2008) Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Molecular Cell* **32**, 1–9.
- Cai, S., Han, H.-J. and Kohwi-Shigematsu, T. (2003) Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nature Genetics* **34**, 42–51.
- Cavalier-Smith, T. (2010) Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. *Biology Direct* **5** (7).
- Ceccarelli, M., Morosi, L. and Cionini, P.G. (1998) Chromocenter association in plant cell nuclei: determinants, functional significance, and evolutionary implications. *Genome* **41**, 96–103.
- Choulet, F., Wicker, T., Rustenholz, C. *et al.* (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. *The Plant Cell Online* **22**, 1686–1701.

- Cohen, T.V., Klarmann, K.D., Sakchaisri, K. *et al.* (2008) The lamin B receptor under transcriptional control of C/EBP ϵ is required for morphological but not functional maturation of neutrophils. *Human Molecular Genetics* **17**, 2921–2933.
- Cokus, S.J., Feng, S., Zhang, X. *et al.* (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219.
- Copenhagen, G.P. and Pikaard, C.S. (1996) RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4. *The Plant Journal* **9**, 259–272.
- D'Ambrosio, C., Schmidt, C.K., Katou, Y., *et al.* (2008) Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. *Genes and Development* **22**, 2215–2227.
- 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 Research* **37**, 3558–3568.
- 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.
- Eils, R., Dietzel, S., Bertin, E. *et al.* (1996) Three-dimensional reconstruction of painted human interphase chromosomes: active and inactive X chromosome territories have similar volumes but differ in shape and surface structure. *Journal of Cell Biology* **135**, 1427–1440.
- Fang, Y. and Spector, D.L. (2005) Centromere positioning and dynamics in living *Arabidopsis* plants. *Molecular Biology of the Cell* **16**, 5710–5718.
- Fedorova, E. and Zink, D. (2008) Nuclear architecture and gene regulation. *Biochimica et Biophysica Acta (BBA) – Molecular Cell Research* **1783**, 2174–2184.
- Feng, S., Jacobsen, S.E. and Reik, W. (2010) Epigenetic reprogramming in plant and animal development. *Science* **330**, 622–627.
- Filion, G.J., van Bommel, J.G., Braunschweig, U. *et al.* (2010) Systematic protein location mapping reveals five principal chromatin types in drosophila cells. *Cell* **143**, 212–224.
- Fischer, A., Hofmann, I., Naumann, K. and Reuter, G. (2006) Heterochromatin proteins and the control of heterochromatic gene silencing in *Arabidopsis*. *Journal of Plant Physiology* **163**, 358–368.
- Fiserova, J., Kiseleva, E. and Goldberg, M.W. (2009) Nuclear envelope and nuclear pore complex structure and organization in tobacco BY-2 cells. *The Plant Journal* **59**, 243–255.
- Fransz, P., De Jong, J.H., Lysak, M. *et al.* (2002) Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proceedings of the National Academy of Sciences of the USA* **99**, 14584–14589.
- Fransz, P. and de Jong, H. (2011) From nucleosome to chromosome: a dynamic organization of genetic information. *The Plant Journal* **66**, 4–17.
- Gaszner, M. and Felsenfeld, G. (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Reviews Genetics* **7**, 703–713.
- Gaudin, V., Libault, M., Pouteau, S. *et al.* (2001) Mutations in LIKE HETEROCHROMATIN PROTEIN 1 affect flowering time and plant architecture in *Arabidopsis*. *Development* **128**, 4847–4858.
- Gaut, B.S. and Ross-Ibarra, J. (2008) Selection on major components of angiosperm genomes. *Science* **320**, 484–486.

- Gerasimova, T.I., Byrd, K. and Corces, V.G. (2000) A chromatin insulator determines the nuclear localization of DNA. *Molecular Cell* **6**, 1025–1035.
- Gindullis, F. and Meier, I. (1999a) Matrix attachment region binding protein MFP1 is localized in discrete domains at the nuclear envelope. *The Plant Cell Online* **11**, 1117–1128.
- Gindullis, F., Pepper, N.J. and Meier, I. (1999b) MAF1, a novel plant protein interacting with matrix attachment region binding protein MFP1, is located at the nuclear envelope. *The Plant Cell Online* **11**, 1755–1768.
- Gindullis, F., Rose, A., Patel, S. and Meier, I. (2002) Four signature motifs define the first class of structurally related large coiled-coil proteins in plants. *BMC Genomics* **3**, 9.
- Grant-Downton, R.T. and Dickinson, H.G. (2005) Epigenetics and its implications for plant biology. 1. The epigenetic network in plants. *Annals of Botany* **96**, 1143–1164.
- Grewal, S.I.S. and Jia, S. (2007) Heterochromatin revisited. *Nature Reviews Genetics* **8**, 35–46.
- Guelen, L., Pagie, L., Brasset, E. *et al.* (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951.
- Happel, N. and Doenecke, D. (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* **431**, 1–12.
- Heitz, E. (1928) Das Heterochromatin der Moose. *Jahrbücher für Wissenschaftliche Botanik* **69**, 762–818.
- Henikoff, S. and Dalal, Y. (2005) Centromeric chromatin: what makes it unique? *Current Opinion in Genetics and Development* **15**, 177–184.
- Heslop-Harrison, J. and Bennett, M. (1990) Nuclear architecture in plants. *Trends in Genetics* **6**, 401–405.
- 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 Research* **9**, 117–121.
- Ingouff, M. and Berger, F. (2009) Histone3 variants in plants. *Chromosoma* **119**, 27–33.
- Jerzmanowski, A. (2007) SWI/SNF chromatin remodeling and linker histones in plants. *Biochimica et Biophysica Acta* **1769**, 330–345.
- Jovtchev, G., Borisova, B.E., Kuhlmann, M. *et al.* (2011) Pairing of lacO tandem repeats in *Arabidopsis thaliana* nuclei requires the presence of hypermethylated, large arrays at two chromosomal positions, but does not depend on H3-lysine-9-dimethylation. *Chromosoma* **120**, 609–619.
- Kankel, M.W., Ramsey, D.E., Stokes, T.L. *et al.* (2003) *Arabidopsis* MET1 Cytosine methyltransferase mutants. *Genetics* **163**, 1109–1122.
- Kanno, T., Bucher, E., Daxinger, L. *et al.* (2008) A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. *Nature Genetics* **40**, 670–675.
- Lahmy, S., Bies-Etheve, N. and Lagrange, T. (2010) Plant-specific multisubunit RNA polymerase in gene silencing. *Epigenetics* **5**, 4–8.
- Lam, E., Luo, C. and Watanabe, N. (2009) Charting functional and physical properties of chromatin in living cells. *Current Opinion in Genetics and Development* **19**, 135–141.
- Lancôt, C., Cheutin, T., Cremer, M. *et al.* (2007) Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature Reviews Genetics* **8**, 104–115.
- Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**, 204–220.

- Lermontova, I., Koroleva, O., Rutten, T. *et al.* (2011) Knockdown of CENH3 in *Arabidopsis* reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. *The Plant Journal* **68**, 40–50.
- Lippman, Z., Gendrel, A.-V., Black, M. *et al.* (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471–476.
- Liu, C., McElver, J., Tzafrir, I. *et al.* (2002) Condensin and cohesin knockouts in *Arabidopsis* exhibit a titan seed phenotype. *The Plant Journal* **29**, 405–415.
- Liu, C., Lu, F., Cui, X. and Cao, X. (2010) Histone methylation in higher plants. *Annual Review of Plant Biology* **61**, 395–420.
- Ma, J., Wing, R.A., Bennetzen, J.L. and Jackson, S.A. (2007) Plant centromere organization: a dynamic structure with conserved functions. *Trends in Genetics* **23**, 134–139.
- Mathieu, O., Jasencakova, Z., Vaillant, I. *et al.* (2003) Changes in 5S rDNA chromatin organization and transcription during heterochromatin establishment in *Arabidopsis*. *Plant Cell* **15**, 2929–2939.
- Mathieu, O., Reinders, J., Caikovski, M. *et al.* (2007) Transgenerational stability of the *Arabidopsis* epigenome is coordinated by CG Methylation. *Cell* **130**, 851–862.
- Mattout, A., Goldberg, M., Tzur, Y. *et al.* (2007) Specific and conserved sequences in *D. melanogaster* and *C. elegans* lamins and histone H2A mediate the attachment of lamins to chromosomes. *Journal of Cell Science* **120**, 77–85.
- Matzke, A.J.M., Huettel, B., van der Winden, J. and Matzke, M. (2005) Use of two-color fluorescence-tagged transgenes to study interphase chromosomes in living plants. *Plant Physiology* **139**, 1586–1596.
- McHale, L., Tan, X., Koehl, P. and Micheltore, R.W. (2006) Plant NBS-LRR proteins: adaptable guards. *Genome Biology* **7**, 212-1–212-11
- Meier, I., Zhou, X., Brkljacic, J. *et al.* (2010) Targeting proteins to the plant nuclear envelope. *Biochemical Society Transactions* **38**, 733.
- Mekhal, K. and Moazed, D. (2010) The nuclear envelope in genome organization, expression and stability. *Nature Reviews Molecular Cell Biology* **11**, 317–328.
- Mirouze, M., Reinders, J., Bucher, E. *et al.* (2009) Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature* **461**, 427–430.
- Moqtaderi, Z., Wang, J., Raha, D. *et al.* (2010) Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nature Structural and Molecular Biology* **17**, 635–640.
- Nakayama, T., Ishii, T., Hotta, T. and Mizuno, K. (2008) Radial microtubule organization by histone H1 on nuclei of cultured tobacco BY-2 cells. *Journal of Biological Chemistry* **283**, 16632–16640.
- Noma, K., Allis, C.D., Grewal, S.I. (2001) Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. *Science* **293**, 1150–1155.
- Noma, K., Cam, H.P., Maraja, R.J. and Grewal, S.I.S. (2006) A role for TFIIC transcription factor complex in genome organization. *Cell* **125**, 859–872.
- Oler, A.J., Alla, R.K., Roberts, D.N. *et al.* (2010) Human RNA polymerase III transcriptsomes and relationships to Pol II promoter chromatin and enhancer-binding factors. *Nature Structural and Molecular Biology* **17**, 620–628.
- Olins, A.L., Rhodes, G., Welch, D.B.M. *et al.* (2010) Lamin B receptor. *Nucleus* **1**, 53–70.
- Pandey, R., Müller, A., Napoli, C.A. *et al.* (2002) Analysis of histone acetyltransferase and histone deacetylase families of *Arabidopsis thaliana* suggests functional diversification of chromatin modification among multicellular eukaryotes. *Nucleic Acids Research* **30**, 5036–5055.

- Paul, A.-L. and Ferl, R.J. (1998) Higher order chromatin structures in maize and *Arabidopsis*. *The Plant Cell Online* **10**, 1349–1360.
- Pecinka, A., Schubert, V., Meister, 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.
- Pecinka, A., Kato, N., Meister, A. *et al.* (2005) Tandem repetitive transgenes and fluorescent chromatin tags alter local interphase chromosome arrangement in *Arabidopsis thaliana*. *Journal of Cell Science* **118**, 3751–3758.
- Pickersgill, H., Kalverda, B., de Wit, E. *et al.* (2006) Characterization of the *Drosophila melanogaster* genome at the nuclear lamina. *Nature Genetics* **38**, 1005–1014.
- Probst, A.V., Fagard, M., Proux, F. *et al.* (2004) *Arabidopsis* histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats. *Plant Cell* **16**, 1021–1034.
- Probst, A.V., Fransz, P.F., Paszkowski, J. and Mittelsten Scheid, O. (2003) Two means of transcriptional reactivation within heterochromatin. *The Plant Journal* **33**, 743–749.
- Rabl, C. (1885) Über Zelltheilung. *Morphol Jahrbuch* 214–330.
- Ravi, M. and Chan, S.W.L. (2010) Haploid plants produced by centromere-mediated genome elimination. *Nature* **464**, 615–618.
- Rosin, F.M., Watanabe, N., Cacas, J.-L. *et al.* (2008) Genome-wide transposon tagging reveals location-dependent effects on transcription and chromatin organization in *Arabidopsis*. *The Plant Journal* **55**, 514–525.
- Rudd, S., Frisch, M., Grote, K. *et al.* (2004) Genome-wide in silico mapping of scaffold/matrix attachment regions in *Arabidopsis* suggests correlation of intragenic scaffold/matrix attachment regions with gene expression. *Plant Physiology* **135**, 715–722.
- Sáez-Vásquez, J. and Gadal, O. (2010) Genome organization and function: a view from yeast and *Arabidopsis*. *Molecular Plant* **3**, 678–690.
- Santos, A.P., Abranches, R., Stoger, E. *et al.* (2002) The architecture of interphase chromosomes and gene positioning are altered by changes in DNA methylation and histone acetylation. *Journal of Cell Science* **115**, 4597–4605.
- Schubert, I. and Shaw, P. (2011) Organization and dynamics of plant interphase chromosomes. *Trends in Plant Science* **16**, 273–281.
- Shibata, F. and Murata, M. (2004) Differential localization of the centromere-specific proteins in the major centromeric satellite of *Arabidopsis thaliana*. *Journal of Cell Science* **117**, 2963–2970.
- Shumaker, D.K., Dechat, T., Kohlmaier, A. *et al.* (2006) Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging. *Proceedings of the National Academy of Sciences* **103**, 8703–8708.
- Soppe, W.J., Jacobsen, S.E., Alonso-Blanco, C. *et al.* (2000) The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Molecular Cell* **6**, 791–802.
- Soppe, W.J.J., Jasencakova, Z., Houben, A. *et al.* (2002) DNA methylation controls histone H3 lysine 9 methylation and heterochromatin assembly in *Arabidopsis*. *EMBO Journal* **21**, 6549–6559.
- Tariq, M., Saze, H., Probst, A.V. *et al.* (2003) Erasure of CpG methylation in *Arabidopsis* alters patterns of histone H3 methylation in heterochromatin. *Proceedings of the National Academy of Sciences* **100**, 8823–8827.

- Tessadori, F., Chupeau, M.-C., Chupeau, Y. *et al.* (2007) Large-scale dissociation and sequential reassembly of pericentric heterochromatin in dedifferentiated *Arabidopsis* cells. *Journal of Cell Science* **120**, 1200–1208.
- Tetko, I.V., Haberer, G., Rudd, S. *et al.* (2006) Spatiotemporal expression control correlates with intragenic scaffold matrix attachment regions (S/MARs) in *Arabidopsis thaliana*. *PLOS Computational Biology* **2**, e21.
- Tikhonov, A.P., Lavie, L., Tatout, C. *et al.* (2001) Target sites for SINE integration in Brassica genomes display nuclear matrix binding activity. *Chromosome Research* **9**, 325–337.
- To, T.K., Kim, J.-M., Matsui, A. *et al.* (2011) *Arabidopsis* HDA6 regulates locus-directed heterochromatin silencing in cooperation with MET1. *PLOS Genetics* **7**, e1002055.
- Towbin, B.D., Meister, P. and Gasser, S.M. (2009) The nuclear envelope – a scaffold for silencing? *Current Opinion in Genetics and Development* **19**, 180–186.
- Tsukahara, S., Kobayashi, A., Kawabe, A. *et al.* (2009) Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature* **461**, 423–426.
- Turck, F., Roudier, F., Farrona, S. *et al.* (2007) *Arabidopsis* TFL2/LHP1 Specifically Associates with Genes Marked by Trimethylation of Histone H3 Lysine 27. *PLOS Genetics* **3**, e86.
- Tutois, S., Cloix, C., Cuvillier, C. *et al.* (1999) Structural analysis and physical mapping of a pericentromeric region of chromosome 5 of *Arabidopsis thaliana*. *Chromosome Research* **7**, 143–156.
- Tzur, Y.B., Wilson, K.L. and Gruenbaum, Y. (2006) SUN-domain proteins: ‘Velcro’ that links the nucleoskeleton to the cytoskeleton. *Nature Reviews Molecular Cell Biology* **7**, 782–788.
- van Driel, R. and Fransz, P. (2004) Nuclear architecture and genome functioning in plants and animals: what can we learn from both? *Experimental Cell Research* **296**, 86–90.
- van Steensel, B., Delrow, J. and Henikoff, S. (2001) Chromatin profiling using targeted DNA adenine methyltransferase. *Nature Genetics* **27**, 304–308.
- Verdaasdonk, J.S. and Bloom, K. (2011) Centromeres: unique chromatin structures that drive chromosome segregation. *Nature Reviews Molecular Cell Biology* **12**, 320–332.
- Vongs, A., Kakutani, T., Martienssen, R.A. and Richards, E.J. (1993) *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**, 1926–1928.
- Wang, G., Zhang, X. and Jin, W. (2009) An overview of plant centromeres. *Journal of Genetics and Genomics* **36**, 529–537.
- Wang, T.-Y., Han, Z.-M., Chai, Y.-R. and Zhang, J.-H. (2010) A mini review of MAR-binding proteins. *Molecular Biology Reports* **37**, 3553–3560.
- Watanabe, K., Pecinka, A., Meister, A. *et al.* (2005) DNA hypomethylation reduces homologous pairing of inserted tandem repeat arrays in somatic nuclei of *Arabidopsis thaliana*. *The Plant Journal* **44**, 531–540.
- Wendt, K.S., Yoshida, K., Itoh, T., Bando, M. *et al.* (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801.
- White, S.A. and Allshire, R.C. (2008) RNAi-mediated chromatin silencing in fission yeast. *Current Topics in Microbiology and Immunology* **320**, 157–183.
- Wicker, T., Sabot, F., Hua-Van, A. *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* **8**, 973–982.
- Widman, N., Jacobsen, S.E. and Pellegrini, M. (2009) Determining the conservation of DNA methylation in *Arabidopsis*. *Epigenetics* **4**, 119–124.

- Wierzbicki, A.T. and Jerzmanowski, A. (2005) Suppression of histone H1 genes in *Arabidopsis* results in heritable developmental defects and stochastic changes in DNA methylation. *Genetics* **169**, 997–1008.
- Wilson, K.L. and Dawson, S.C. (2011) Functional evolution of nuclear structure. *The Journal of Cell Biology* **195**, 171–181.
- Wood, A.J., Severson, A.F. and Meyer, B.J. (2010) Condensin and cohesin complexity: the expanding repertoire of functions. *Nature Reviews Genetics* **11**, 391–404.
- Wyrick, J.J. and Parra, M.A. (2009) The role of histone H2A and H2B post-translational modifications in transcription: A genomic perspective. *Biochimica et Biophysica Acta* **1789**, 37–44.
- Yusufzai, T.M. and Felsenfeld, G. (2004) The 5'-HS4 chicken β -globin insulator is a CTCF-dependent nuclear matrix-associated element. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 8620–8624.
- Zellinger, B. and Riha, K. (2007) Composition of plant telomeres. *Biochimica et Biophysica Acta* **1769**, 399–409.
- Zhang, H. and Dawe, R.K. (2011) Mechanisms of plant spindle formation. *Chromosome Research* **19**, 335–344.
- Zhou, X., Graumann, K., Evans, D.E. and Meier, I. (2012) Novel plant SUN-KASH bridges are involved in RanGAP anchoring and nuclear shape determination. *The Journal of Cell Biology* **196**, 203–211.
- Zilberman, D., Coleman-Derr, D., Ballinger, T. and Henikoff, S. (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**, 125–129.