



HAL
open science

Transmission des marques épigénétiques lors de la multiplication sexuée et de la propagation asexuée chez le pommier (*Malus domestica*)

Adrien Perrin

► To cite this version:

Adrien Perrin. Transmission des marques épigénétiques lors de la multiplication sexuée et de la propagation asexuée chez le pommier (*Malus domestica*). Sciences agricoles. Université d'Angers, 2020. Français. ⟨NNT : 2020ANGE0010⟩. ⟨tel-03935853⟩

HAL Id: tel-03935853

<https://theses.hal.science/tel-03935853v1>

Submitted on 12 Jan 2023

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.



HAL Authorization

Thèse de doctorat de

L'UNIVERSITE D'ANGERS
Comue Université Bretagne Loire

ECOLE DOCTORALE N° 600
Ecole doctorale Ecologie, Géosciences, Agronomie et Alimentation
Spécialité : Biologie et physiologie végétales

Par

Adrien PERRIN

Transmission des marques épigénétiques lors de la multiplication sexuée et de la propagation asexuée chez le pommier (*Malus domestica*)

Thèse présentée et soutenue à Angers, le 15 mai 2020

Unité de recherche : IRHS

Thèse N° :

Rapporteurs avant soutenance :

Annabelle DEJARDIN - Chargée de recherche - INRAe
Stéphane MAURY – Pr. des universités - Université d'Orléans

Composition du Jury :

Présidente : Nathalie LEDUC – Pr. des universités - Université d'Angers
Annabelle DEJARDIN - Chargée de recherche – INRAe (Orléans)
Fabrice FOUCHER – directeur de recherche – INRAe (Angers)
Stéphane MAURY – Pr. des universités - Université d'Orléans
Clémentine VITTE – Chercheuse – CNRS – GQE- Le Moulon (Gif-sur-Yvette)

Directeur de thèse

Etienne BUCHER – Chercheur – Agroscope (Suisse)

Co-encadrant

Jean-Marc CELTON – Maître de conf. – Université d'Angers

Emilie VERGNE – Ingénieur de recherche - INRAe

Si tu as une pomme, que j'ai une pomme, et que l'on échange nos pommes, nous aurons chacun une pomme. Mais si tu as une idée, que j'ai une idée et que l'on échange nos idées, nous aurons chacun deux idées.

George Bernard Shaw

L'auteur du présent document vous autorise à le partager, reproduire, distribuer et communiquer selon les conditions suivantes :



- Vous devez le citer en l'attribuant de la manière indiquée par l'auteur (mais pas d'une manière qui suggérerait qu'il approuve votre utilisation de l'œuvre).
- Vous n'avez pas le droit d'utiliser ce document à des fins commerciales.
- Vous n'avez pas le droit de le modifier, de le transformer ou de l'adapter.

Consulter la licence creative commons complète en français :
<http://creativecommons.org/licences/by-nc-nd/2.0/fr/>

Ces conditions d'utilisation (attribution, pas d'utilisation commerciale, pas de modification) sont symbolisées par les icônes positionnées en pied de page.



REMERCIEMENTS

Une thèse de doctorat est un travail de recherche et d'apprentissage par la recherche long et exigeant. Réaliser une thèse m'a permis d'apprendre sur moi-même et sur ce que je souhaite pour mon futur. J'ai également appris à travailler en communauté et à collaborer avec d'autres personnes pour atteindre un objectif. Ce travail n'aurait pas pu aboutir sans l'aide et les conseils de nombreuses personnes et j'aimerais remercier plus particulièrement certaines d'entre elles ici.

Je remercie mon directeur de thèse, Mr Etienne BUCHER qui m'a donné la possibilité de réaliser une thèse passionnante. J'ai beaucoup appris à ses côtés.

Je remercie également mes deux co-encadrants, Mme Emilie VERGNE et Mr Jean-Marc CELTON.

Jean-Marc pour ta bonne humeur, ton aide précieuse notamment pendant la rédaction et tes questions acerbes qui m'ont agacées plus d'une fois mais qui étaient toujours tellement pertinentes et qui m'ont permises de prendre du recul.

Emilie, je te remercie pour ta disponibilité et ta bienveillance. Tu as toujours été là quand j'avais besoin d'aide et quand j'avais besoin de discuter pour lever mes doutes. Merci aussi pour les discussions autour de notre passion commune: la couture.

Sandrine, merci pour toute ton aide, pour tes questions qui m'ont fait remettre mon travail en perspective et ton soutien. Et merci aussi, parce que « On en a Gros » !

Elisabeth et Nicolas Dousset, merci pour toute votre aide et votre bienveillance. J'ai appris beaucoup de chose sur la transformation et la culture *in vitro* grâce à vous.

Je remercie également Béatrice pour sa gentillesse et sa persévérance malgré nos déboires sur certaines transformations et sur les cultures *in vitro*.

Merci aussi à l'équipe bio-info et plus particulièrement à Claudine, Sylvain, Sébastien et Nicolas Daccord Pour leur soutien en bio-informatique sans lequel ce travail aurait pris beaucoup plus de temps.

Merci à tous mes collègues et toutes les personnes avec qui j'ai travaillé pendant cette thèse, notamment Christophe, Aurélia, Fabienne et Murielle, mais également Michel et Christian.

Merci aux doctorants, parce que la petite bière une fois par mois avec vous devrait être remboursé par la sécu tellement ça redonne le moral.

David R., je tiens d'abord à te remercier pour toute ton aide en tant que collègue et maintenant pour ton amitié.

Je ne t'oublie pas non plus, mon petit Kay, merci pour ton amitié et ton petit côté allemand. Ich liebe dich mein freund.

Merci aux « Copains Angevins », pour les repas, les gouter, les apéros et les weekends découvertes en Normandie et en Bretagne.

Merci à ma famille et ma belle-famille pour votre soutien.

Je pense aussi à mon chat (« el diablo » d'après certain) pour la ronronthérapie et son soutien nonchalant.

Enfin je remercie ma femme. Merci pour ton aide de relecture qui a rendu mon manuscrit plus intelligible qu'il ne l'était. Merci aussi, parce que l'on a tous deux vécu trois années longues et difficiles, malgré lesquels nous avons continué à construire notre avenir ensemble. Merci, parce que tu m'as toujours soutenue et poussée à continuer dans les moments difficiles ou j'aurais pu abandonner.

A mon père

Sommaire

ABREVIATION	1
INTRODUCTION GENERALE	3
1. Malus domestica Borkh	5
1.1. Origine	5
1.2. Production mondiale et Française	5
1.3. Biologie.....	5
2. Multiplications sexuée et asexuée	7
2.1. La reproduction sexuée	7
2.1.1. Cas du pommier	7
2.2. La multiplication asexuée.....	9
2.2.1. Origine du greffage	9
2.2.2. Fonction et mise en place de la greffe	9
2.2.3. Epigénétique et greffage.....	10
2.2.4. Intérêt du greffage chez le pommier	11
3. La génétique chez les plantes	13
3.1. Le génome des plantes	13
3.1.1. Le génome du pommier	15
a) Les versions du génome de GDDH13	16
3.2. Transcription de l'ADN	17
3.2.1. Les ARNs messagers ou ARNs codants	17
3.2.2. Les ARNs non-codants	17
3.3. Traduction de l'ARNm	18
3.4. Régulation de l'expression des gènes	18
3.4.1. Régulation génétique.....	19
3.4.2. Régulation post-transcriptionnelle.....	19
4. Régulation épigénétique chez les plantes	21
4.1. Définition de l'épigénétique	21
4.2. Les histones	21
4.2.1. Définition	21
4.2.2. Modification des histones	21
a) L'acétylation de la queue histone	21
b) La méthylation de la queue histone.....	22
c) Acétylation et méthylation des histones chez le pommier	22
4.3. La méthylation de l'ADN	23
4.3.1. Définition	23
a) Chez le pommier.....	23
4.3.2. Rôle de la méthylation de l'ADN.....	24
a) Méthylation de l'ADN et éléments transposables.....	24
4.3.3. Mécanisme de méthylation de l'ADN	27
a) La maintenance de la méthylation.....	27
b) La méthylation de novo	29
4.3.4. Méthylation de l'ADN et différences entre plantes pérennes et annuelles	29
4.3.5. Méthylation de l'ADN et amélioration des plantes	30
5. Transmission de l'information épigénétique	31
5.1. Multiplication asexuée	31
5.1.1. Cas du greffage	31
5.1.2. Cas de la multiplication in vitro	31
5.2. Multiplication sexuée	32
5.2.1. Gamétogénèse mâle.....	32
5.2.1. Gamétogénèse femelle	32
5.2.2. Fécondation.....	33
6. Objet de la thèse	35
6.1. Contexte	35
6.2. Objectif.....	35

CHAPITRE I : DIFFERENCES EPIGENETIQUES ENTRE UNE MULTIPLICATION SEXUEE ET UNE PROPAGATION ASEXUEE CHEZ LE POMMIER 37

1.	Introduction	39
2.	Matériels et méthodes spécifiques	41
2.1.	Matériel végétal	41
2.2.	Séquençage bisulfite	41
3.	Partie 1 – "Divergent DNA methylation signatures of juvenile seedlings, grafts and adult apple trees" (Perrin et al. 2020)	43
3.1.	Introduction à l'article	43
3.2.	Divergent DNA methylation signatures of juvenile seedlings, grafts and adult apple trees	43
3.3.	Résumé de l'article.....	64
4.	Partie 2 - Résultats complémentaires	65
4.1.	Matériels et méthodes	65
4.1.1.	Matériel végétal	65
4.1.2.	Fixation et coupes histologiques des échantillons YG et SD.....	65
4.1.3.	Microarray et calcul des régions différenciellement méthylées	66
4.2.	Résultats.....	66
4.2.1.	Analyse de la troisième comparaison : plants greffés versus plants semés	66
a)	Phénotypage	66
b)	Transcrits différenciellement exprimés	73
c)	Régions différenciellement méthylées	77
d)	Régions différenciellement méthylées en proximité des gènes.....	79
4.2.2.	Comparaison de deux méthodes de calculs utilisées pour l'analyse des DMRs	80
a)	Comparaison globale des deux méthodes de calculs	80
b)	Première méthode : Découpage en régions autour des gènes	82
c)	Seconde méthode : Sliding windows 200pb	85
4.3.	Discussion	86
4.3.1.	Effets de la transmission sexuée ou asexuée des marques épigénétiques	86
a)	Effets phénotypiques	86
b)	Effet sur le niveau de transcription.....	88
c)	Régions différenciellement méthylées	88
d)	Est-ce que l'analyse de la comparaison YGvsSD corrobore les observations faites précédemment pour SDvsOG et OGvsYG ?	89
4.3.2.	Discussion des méthodes de calculs des DMRs utilisées et proposition d'autres méthodes de calculs	90
a)	Était-il judicieux de faire un calcul sur le génome complet plutôt que sur des régions spécifiques ?	90
b)	Pourquoi avoir choisi de faire les calculs avec la deuxième méthode plutôt que la première ?	91
c)	Quelles autres méthodes de calcul aurions-nous pu mettre en place et quelles seraient leurs avantages	92
d)	Quelle méthode choisir ?.....	97
4.3.3.	Critiques de différent points	98
a)	L'utilisation de la classification par code BIN : une bonne idée ?.....	98
b)	Discussion du plan expérimental	98
5.	Conclusion	101

CHAPITRE II – EDITION DU GENOME ET DE L'EPIGENOME DU POMMIER 103

1. Introduction	105
1.1. La technique CRISPR-Cas9	105
1.1.1. Composition du système CRISPR-Cas9 et modalités de réparation des cassures double brin de l'ADN	105
1.1.2. Evolution des applications de l'édition du génome	107
1.1.3. Etat de l'art des technologies CRISPR-Cas9 chez le pommier	108
1.2. Mutagenèse dirigée du gène MET1	108
1.3. Modification de l'épigénome du promoteur du gène TFL1	108
2. Tentative d'extinction d'une voie de maintenance de la méthylation chez le pommier par édition des gènes codant pour la fonction MET1	111
2.1. Matériel et méthodes	111
2.2. Résultats.....	115
2.2.1. Recherche des orthologues de AtMet1 chez le pommier	115
a) V1.0 du génome de GDDH13.....	115
b) V1.1 du génome de GDDH13.....	117
c) Chez le génotype 'Gala'	118
2.2.2. Mise en place de la technique CRISPR-Cas9 pour réaliser l'extinction des gènes Md02MET1 et Md15MET1 du pommier	119
a) Sélection des crRNAs pour le génotype GDDH13	119
b) Sélection des crRNAs pour le génotype 'Gala'	120
2.2.3. Résultats des transformations de vitroplants de pommiers	120
2.2.4. Analyse moléculaire des lignées transformées	122
2.2.5. Phénotypage en serre des lignées mutantes	124
2.2.6. Caractérisation épigénétique des mutants	126
a) Epigénotypage	126
b) Séquençage bisulfite du génome complet	129
2.3. Discussion	129
2.3.1. Efficacité de transformation	129
a) Tentative de transformation du génotype GDDH13	129
2.3.2. Un taux important de profil d'édition homozygote	130
2.3.3. La lignée 2P un hétérozygote pour toujours ?	131
2.3.4. Les lignées 2C2 et 2C3 sont peut-être une seule et même lignée	131
2.3.5. Séquençage bisulfite et test de digestion	131
2.3.6. Phénotypes attendus pour des plantes dont les voies de maintien de la méthylation sont défectueuses	132
3. Edition de l'épigénome du pommier – modification de la méthylation d'un gène pour modifier son expression	135
3.1. Choix méthodologiques en vue de l'obtention d'une construction permettant la méthylation d'une séquence cible	135
3.1.1 Choix des crRNA ciblant le promoteur de MdTFL1	135
3.1.2. Choix de la méthyltransférase et de son promoteur.....	135
3.1.3 Choix et obtention de la Cas9 inactivée.....	137
a) Premier protocole : mutation directement sur le plasmide pDeCas9-Kr (Figure 3.1-3 et Annexe 9).....	139
b) Second protocole : mutation par amplifications et clonages successifs à partir du plasmide pDeCas9-Kr (Figure 3.1-4, Annexe 10)	140
c) Troisième protocole : utilisation d'une deadCas9 d'un plasmide existant (Figure 3.1-5).....	142
3.1.2. Obtention de la portion HSP::DRM2 (Figure 3.1-1E)	143
3.1.3. Obtention d'un plasmide contenant la construction HSP::DRM2::deadCas9 à partir, ou non, de pDe-deadCas9-Kr.....	143
a) Protocole d'obtention par Golden Gate (Figure 3.1-6)	143
b) Protocole d'obtention par clonage conventionnel (Figure 3.1-6)	144
c) Protocole d'obtention à l'aide d'un plasmide intermédiaire (Figure 3.1-7).....	145
d) Protocole d'insertion de la construction HSP::DRM2::deadCas9 synthétisée (Figure 3.1-8)	146
3.2. Résultats et discussion	147
3.2.1. Obtention des crRNAs ciblant le promoteur de MdTFL1	147

3.2.2.	<i>Obtention d'une deadCas9 fonctionnelle</i>	149
a)	Premier protocole : mutation directement sur le plasmide pDeCas9-Kr (Figure 3.1-3).....	149
b)	Second protocole : mutation par amplifications et clonages successifs à partir du plasmide pDeCas9-Kr (Figure 3.1-4).....	149
c)	Troisième protocole : Utilisation d'une deadCas9 d'un plasmide existant (Figure 3.1-5).....	150
3.2.3.	<i>Echec de l'obtention de la construction HSP::DRM2 ::deadCas9 dans le plasmide final</i>	150
a)	Choix de la méthyltransférase DRM2	150
b)	Obtention de la construction HSP::DRM2 ::deadCas9 par Golden Gate (Figure 3.1-6).....	150
c)	Obtention de la construction HSP::DRM2 ::deadCas9 par clonage conventionnel (Figure 3.1-6).....	151
d)	Protocole d'obtention à l'aide d'un plasmide intermédiaire (Figure 3.1-7).....	151
e)	Protocole d'insertion de la construction HSP::DRM2::deadCas9 synthétisée (Figure 3.1-8).....	151
3.2.4.	<i>Proposition d'un protocole amélioré pour l'obtention de la construction HSP::DRM2 ::deadCas9 dans le plasmide final</i>	152
a)	Synthèse complète	152
b)	Protocole à mettre en place pour obtenir le plasmide pDE-HSP-DRM2-deadCas9-kr	152
4.	<i>Conclusion – Chapitre II</i>	155
	CONCLUSION GENERALE	157
	BIBLIOGRAPHIE	163
	ANNEXES	183
	Annexes à l'article (figures supplémentaires)	183
	Annexes Chapitre I	190
	Annexes Chapitre II	196
	TABLE DES ILLUSTRATIONS	219
	TABLE DES TABLEAUX	221

Abréviation

Malus domestica Borkh : *M. domestica*

n.b : nota bene

p.e. : par exemple

cf : confer

DET : Transcrit différenciellement exprimé

DMR : Région différenciellement méthylé

DEG : Gène différenciellement transcrit

DETE : Elément transposable différenciellement transcrit

DEG-DMR : Gene différenciellement transcrit proche d'une région différenciellement méthylé

DETE-DMR : Elément transposable différenciellement transcrit proche d'une région différenciellement méthylé.

SDvsOG : comparaison entre un pommier adulte et un jeune pommier issu de semis

OGvsYG : comparaison entre un pommier adulte et un jeune pommier greffé

YGvsSD : comparaison entre un pommier greffé et un pommier semée la même année.

GDDH13 : Golden delicious double haploid n°13

PCR : Polymerase chain reaction – réaction d'amplification de l'ADN

MET1 : Méthyltransférase 1

TFL1 : terminal flowering locus 1

Pb : paire de base (une base = un nucléotide composant la séquence de l'ADN)

Avant-propos : l'introduction à ma thèse proposée ci-après se veut accessible pour les néophytes en science et plus particulièrement en biologie végétale et biologie moléculaire. Certaines notions qui y sont introduites pourront donc paraître non-nécessaire ou évidente aux lecteurs avertis.

Introduction générale

Pour ce travail de thèse, il m'a été confié la responsabilité de découvrir une partie des secrets d'un fruit mythique souvent associé, à tort, à la suite d'une mauvaise interprétation de sa traduction latine, au fruit défendu recelant toute la connaissance de l'univers : La Pomme.

La pomme fut utilisée dans l'histoire pour embellir des découvertes tel que l'énonciation des principes de la gravité par Isaac Newton (non, il ne l'a pas reçu sur la tête) ou des citations telle que la citation «An apple a day keeps the doctor away, if you aim it well enough » attribuée à Winston Churchill (finalement ce n'est pas Winston Churchill qui l'a prononcée, communication personnelle du Professeur Freeman, International Churchill Society). La pomme fait aussi partie de l'imaginaire collectif comme un fruit mythique et mystique (www.lapomme.org) et ce depuis, l'enfance en passant par la pomme empoisonnée dans laquelle croque Blanche-neige ou encore son utilisation dans la comptine « pomme de Reinette et pomme d'Api ».

Je me suis donc intéressé à la pomme, ou plutôt à l'arbre sur lequel elle est produite, le pommier et plus particulièrement à son bagage épigénétique.

1. *Malus domestica* Borkh

1.1. Origine

Il existe de nombreuses espèces de pommier dans la nature. L'espèce la plus cultivée dans le monde est *Malus domestica* (*M. domestica*; Gross *et al.*, (2014)) décrite par Borkh en 1803. L'espèce *M. domestica* Bork est apparue il y a entre 4.000 et 10.000 ans en Asie centrale (Cornille *et al.*, 2012). Elle a divergé de l'espèce sauvage *Malus sieversii*, originaire d'Asie centrale et qui a été identifiée comme étant la contributrice majoritaire au patrimoine génétique de *M. domestica* (Robinson *et al.*, 2001; Harris *et al.*, 2002). Une série d'hybridations (*n.b.* croisement entre deux espèces) a ensuite eu lieu avec *M. orientalis* et *M. sylvestris* (Coart *et al.*, 2006; Cornille *et al.*, 2014). Ces hybridations ont été favorisées par l'introduction de *M. domestica* en Europe par les Romains (Schlumbaum *et al.*, 2012) et par les échanges commerciaux le long de la route de la soie (Cornille *et al.*, 2012).

1.2. Production mondiale et Française

La pomme est le troisième fruit le plus produit au monde avec plus de 80 millions de tonnes produites en 2017 (FAOSTAT 2017). En France ce fruit représente la deuxième plus importante production derrière le raisin.

En 2019 pas moins de 22 variétés de pommes à couteau ont été produites en France pour la commercialisation. Parmi les plus cultivées on retrouve la 'Golden Delicious', la 'Gala', la 'Pink Lady' ou encore la 'Granny Smith' (Poires Association Nationale Pomme, 2019). Il existe également des variétés de pommes à cidre (Les Mordus de la Pomme, 2020).

1.3. Biologie

Le pommier appartient à la sous-famille des Maloideae au sein de la famille des Rosaceae. *M. domestica* est une espèce utilisée pour la production dans des habitats tempérés (Cornille *et al.*, 2012). Ce sont des arbres ou arbustes à feuillage caduques présentant une phyllotaxie alterne. Les fleurs sont regroupées en corymbe de 5 à 6 fleurs. La floraison a lieu habituellement entre avril et mai dans l'hémisphère nord.

Le pommier a une durée de vie de cinquante ans et plus (Canadian Food Inspection Agency, 2020). En pratique, dans les vergers, les pommiers sont remplacés en moyenne tous les 20 ans, car leur production diminue au-delà.

Le pommier possède une période de juvénilité pouvant se prolonger au-delà de 8 ans (Visser, 1964).

2. Multiplications sexuée et asexuée

Chez tous les êtres vivants la reproduction est une étape importante de la vie car elle permet la pérennisation d'une espèce. Chez les végétaux la reproduction peut prendre deux formes : soit la reproduction est sexuée, soit la reproduction est asexuée.

2.1. La reproduction sexuée

La reproduction sexuée fait intervenir les organes reproducteurs des plantes et consiste en la fécondation d'un gamétophyte femelle mature par un grain de pollen mature (Kawashima & Berger, 2014). De cette fécondation résulte la production d'un embryon à partir duquel se développera l'organisme nouvellement formé. La reproduction sexuée permet un brassage génétique résultant en la formation d'un organisme partageant une partie des caractères génétique des gamétophytes femelle et mâle.

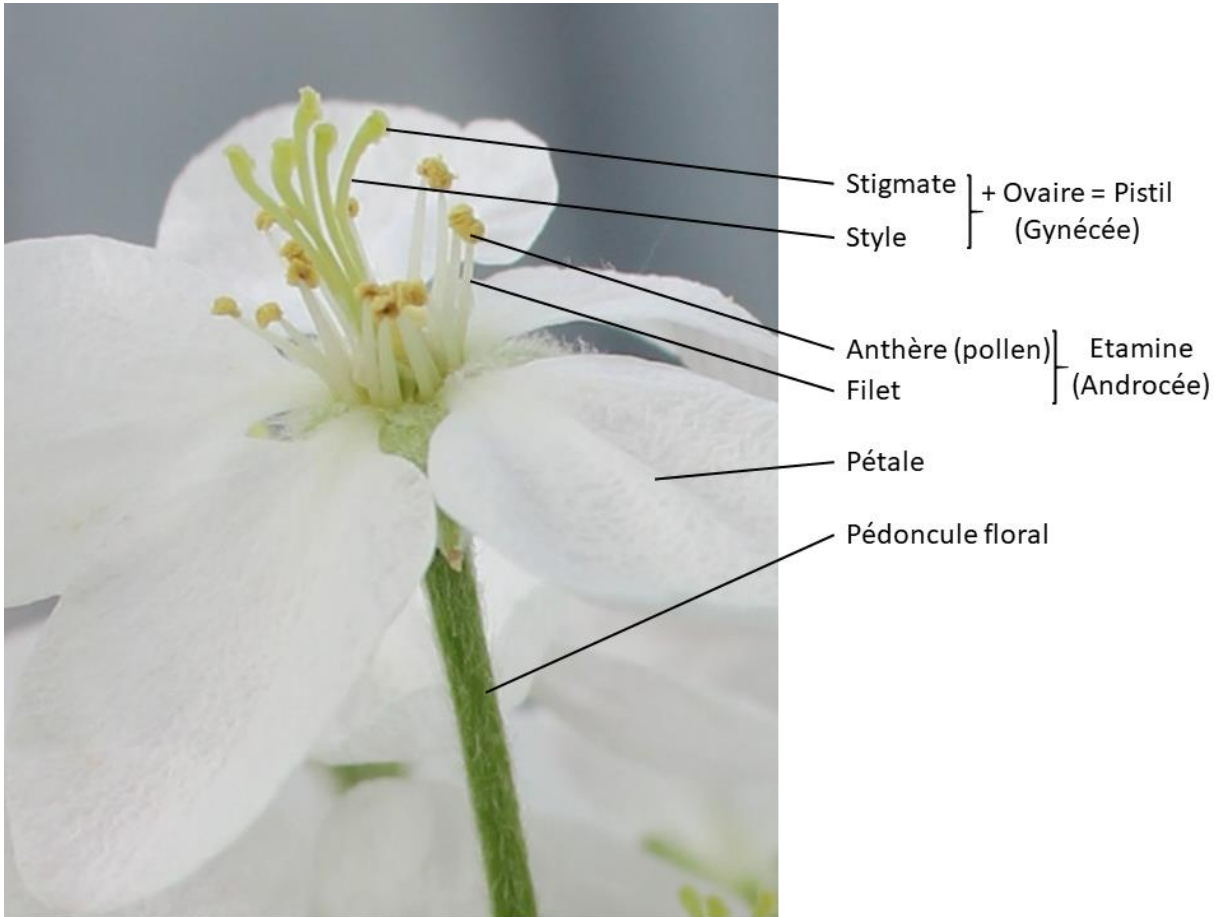
2.1.1. Cas du pommier

Bien que présentant les organes reproducteurs mâles et femelles au sein de la même fleur (Figure 2.1-1A), le pommier présente une auto-incompatibilité pollinique (Zhang *et al.*, 1988). Lors de la pollinisation, le grain de pollen se colle au stigmate. Le pollen va ensuite germer et former un tube pollinique qui passera dans le style pour atteindre l'ovaire et l'ovule qu'il contient, afin de réaliser la fécondation. Au sein du style des molécules de reconnaissance sont produites (revu dans Orcheski and Brown (2012)). Ces molécules permettent à l'arbre de reconnaître si le pollen provient de ses propres fleurs ou de celles d'un arbre différent. Pour chaque génotype de pommier, les molécules de reconnaissance sont produites à partir de deux allèles d'un gène appelé gène S. Chaque allèle est identifié par un numéro (e.g. S₁, S₂, etc...) et il existe environ 28 allèles différents. Chaque génotype de pommier possède donc une combinaison spécifique d'allèles S (e.g. pour la variété 'Golden Delicious' : S₂S₃). Lorsque le tube pollinique passe dans le style il y a donc reconnaissance des allèles S : trois cas sont alors possible (Figure 2.1-1B). Soit une incompatibilité complète lorsque le pollen possède les mêmes allèles S que la fleur pollinisée. Soit une compatibilité partielle lorsque l'un des allèles est différent des allèles de la fleur pollinisée. Et enfin une compatibilité parfaite lorsque les deux allèles sont différents de ceux de la fleur pollinisée (Orcheski & Brown, 2012).

Si le pollen provient du même arbre, de la même variété ou d'une variété proche, la pollinisation sera donc bloquée. Ceci signifie également que dans la grande majorité des cas, chaque pépin au sein d'une pomme représente un futur individu portant un mélange des caractères des deux génotypes parentaux, et n'est donc pas une « copie » de l'un ou l'autre des parents.

On notera par ailleurs que si une autofécondation était possible, le caractère hautement hétérozygote du pommier rendrait la possibilité d'obtenir un cultivar similaire hautement improbable, notamment du au brassage génétique lors de la méiose, puis de la fécondation

A.



B.

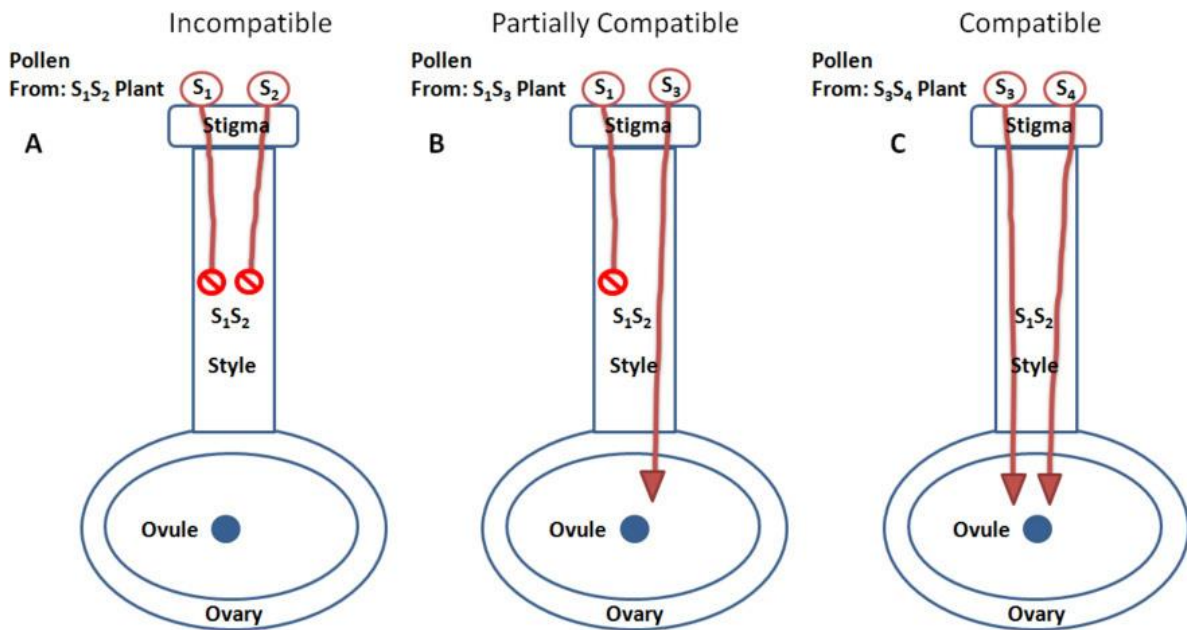


Figure 2.1-1 : Anatomie d'une fleur de pommier en A. (Photo : A. Perrin) et diagramme de pollinisation en B. (Orcheski et al 2012).

2.2. La multiplication asexuée

La multiplication asexuée ou multiplication végétative est un mode de reproduction commun chez les végétaux (exemples : fraisiers, iris, chlorophytum, etc...). Il permet à une plante de produire un nouvel individu à partir d'une tige ou d'une racine. L'individu nouvellement formé possédera alors le même patrimoine génétique que son parent, ce sont des clones.

Pour ma thèse je me suis intéressé à un mode particulier de multiplication asexuée : la propagation asexuée par greffage. Le greffage est un processus spécifique au dicotylédone. En effet l'absence de cambium chez les monocotylédone ne permet pas ce processus (Gautier *et al.*, 2019).

2.2.1. Origine du greffage

Bien que des tablettes en écriture cunéiforme datant de près de 2000 ans avant J.C. faisaient déjà mention du greffage Mudge *et al.* (2009), ce n'est pas une invention de l'Homme, bien qu'il ait largement développé cette technique. En effet deux plantes voisines ou deux branches d'un même arbre sont capables de fusionner naturellement leur système vasculaire comme l'a montré Mudge *et al.* (2009) dans la très complète revue *A History of Grafting*.

2.2.2. Fonction et mise en place de la greffe

Au sein d'une plante greffée on définit deux parties : le scion ou greffon qui est la partie aérienne de la plante, et le porte greffe (ou « rootstock » en anglais) qui porte le scion et présente le système racinaire de la plante greffée.

L'intérêt du greffage est que chaque partie de la plante greffée va influencer le développement de l'autre partie en fonction de ses caractéristiques intrinsèques. En sylviculture, on recherche plutôt à orienter l'effet du porte greffe vers le scion. C'est-à-dire que l'effet du porte greffe sur le scion doit être plus important que l'effet du scion sur le porte greffe (Mudge *et al.*, 2009).

Le greffage permet notamment d'améliorer la résistance aux agents pathogènes. Ainsi chez la vigne, le greffage permet d'améliorer la résistance au phylloxera (Mudge *et al.*, 2009; Kumari *et al.*, 2015).

Malgré un intérêt majeur notamment pour la culture, chez les plantes pérennes ou potagères (Rouphael *et al.*, 2018), la compréhension des mécanismes sous-jacents à la greffe restent mal-connus (Melnik, 2017a,b; Pina *et al.*, 2017; Berger *et al.*, 2018). La revue réalisée par (Nanda & Melnik, 2018) explique notamment le rôle des hormones dans la fusion des deux systèmes vasculaires.

Le greffage consiste donc en l'union de deux tissus adultes présentant des cellules différenciées (revu dans (Berger *et al.*, 2018)). La mise en place de la greffe nécessite un processus de dédifférenciation des tissus et de formation d'un nouveau système conducteur. On peut ainsi différencier plusieurs étapes de la greffe. Il y a tout d'abord un processus d'adhésion entre le scion et le porte greffe. Ceci est rendu possible par la formation d'une interface riche en pectine (Berger *et al.*, 2018). Il y a ensuite formation d'un cal cellulaire au point de jonction. La couche pectinique va ensuite disparaître. On peut observer en parallèle la formation de plasmodesmes permettant des échanges d'organites (Berger *et al.*, 2018; Gautier *et al.*, 2019). On notera également qu'au niveau du point de greffe, les cellules formant un cal peuvent présenter une polyploïdisation. Ce qui a notamment été montré par (Fuentes *et al.*, 2014) chez le tabac. Les cellules présentaient un génome nucléaire et chloroplastique issus de la fusion du génome du scion et du porte greffe.

Enfin le cambium formé au point de contact de la greffe va se différencier et permettre la formation d'un nouveau système vasculaire (Berger *et al.*, 2018). Chez *A.thaliana*, la mise en place du système vasculaire apparait entre 3 et 7 jours après l'initiation de la greffe (Melnik & Meyerowitz, 2015).

Chez *A.thaliana*, il a été montré que des plantes greffées présentent une modification de l'expression des gènes. Dans ces plantes, il y a une augmentation de l'expression des gènes liés à la formation des tissus vasculaires (Melnik *et al.*, 2018)

Chez les plantes pérennes le greffage est réalisé sur du bois dormant. Le succès de la greffe n'est mesurable que lors du débourrement du bourgeon et les tissus concernés sont difficiles à analyser. Ces caractéristiques rendent plus difficile l'étude de la greffe chez les plantes pérennes . Cependant des études ont été réalisées sur la vigne, notamment via des techniques d'imagerie à résonance magnétique ou de rayon X (Bahar *et al.*, 2010; Milien *et al.*, 2012) afin d'étudier le point de greffe. Ces études ont mis en évidence l'importance de la mise en place d'un réseau vasculaire bien développé pour le succès de la greffe.

2.2.3. Epigénétique et greffage

Berger *et al.*, (2018) propose une revue des récentes découvertes sur les processus épigénétiques mis en œuvre par les plantes. Il présente notamment un dialogue croisé important entre hormone, modification épigénétique et modification de l'expression de gènes. Il semble que des modifications épigénétiques soient requises notamment pour la formation de cal, ce qui a été montré *in vitro*. Berger *et al.*, (2018) présentent également les liens entre formation d'un cal lors d'une blessure et modifications épigénétiques. Enfin ils présentent les modifications épigénétique associées à la différenciation cellulaire.

Berger *et al.*, (2018) font état du manque de connaissance sur la régulation épigénétique lors du greffage et propose un modèle intégrant les connaissances des différents processus connus à ce jours, qui pourraient entrer en jeu lors de la greffe et permettent la jonction et la communication entre scion et porte greffe (Figure 2.2-1).

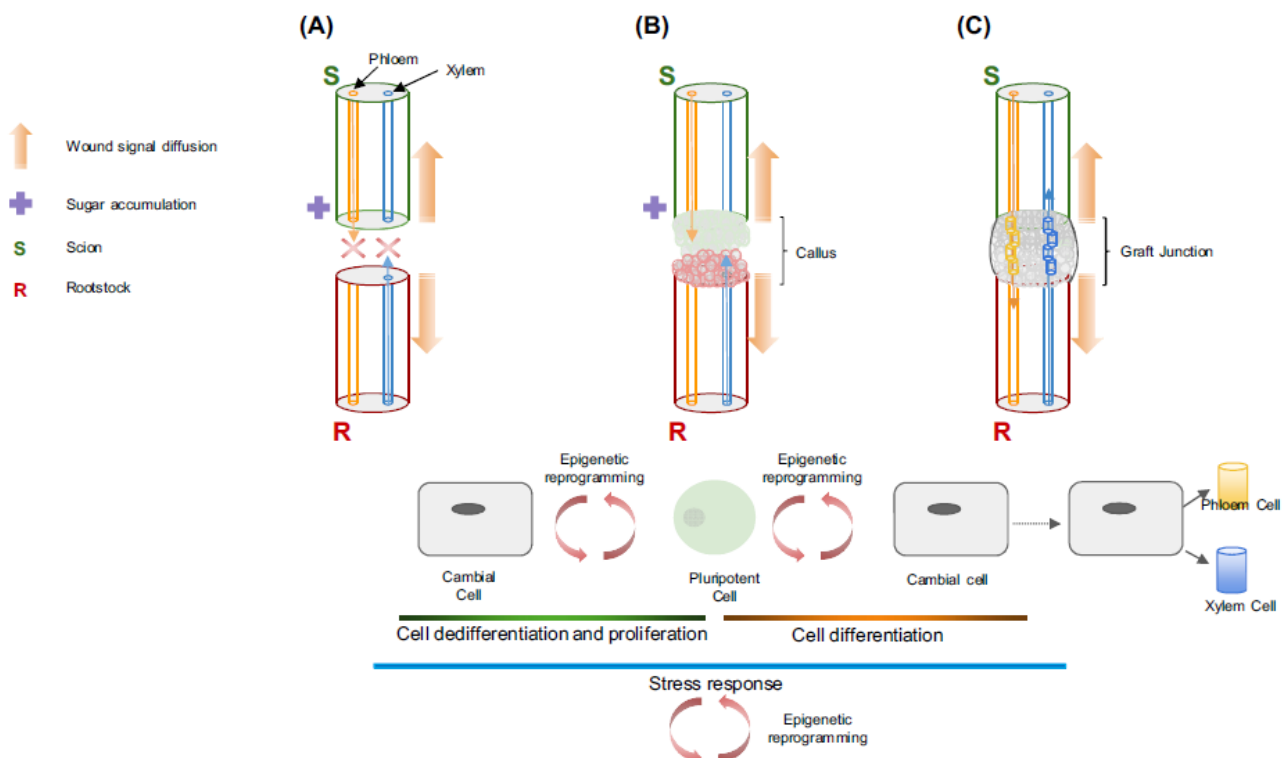


Figure 2.2-1: modèle proposé par Berger et al 2018: " Figure 2: Early events during graft formation involve epigenetic remodeling. (A) Initial situation. Wounding and loss of plant integrity generate a stress response in the grafted plant, locally at the graft junction, and also potentially to the entire plant through long distance signaling. Hormonal and metabolic fluxes are interrupted, leading to an aberrant physiological situation at the graft interface as illustrated by sugar accumulation in the scion but not in the rootstock. Because epigenetic modifications are components of the stress response, grafting most probably induce stress-related epigenetic reprogramming. (B) Callogenesis. One of the earliest events during grafting is the formation of a callus between the scion and the rootstock: cells mainly originating from the cambium reverse to a less differentiated, pluripotent state and regain their competence for proliferation. An extensive number of data show that cell dedifferentiation requires transcriptional reprogramming depending on epigenetic remodeling. (C) Vascular reconnection. Vascular reconnection is considered as the basic requirement for a successful grafting. It relies on the formation of a new cambium and the production of secondary xylem and phloem from this cambium. Although the involvement of epigenetic regulations in cambium initiation and functioning has not been shown yet, different data point to a general role for epigenetic mechanism during cell differentiation".

2.2.4. Intérêt du greffage chez le pommier

Précédemment, nous avons vu qu'une plante greffée est une chimère de deux plantes. Chaque partie conserve son patrimoine génétique propre. Le porte greffe peut apporter différents bénéfices pour la sylviculture (revu dans Singh *et al.* 2019), tel qu'une diminution de la période de juvénilité, un contrôle de la taille de l'arbre (Simons, 1986), de son port (Jensen *et al.*, 2010), ou une meilleure adaptation à un terroir (Ahmed *et al.*, 2013). Le greffage chez le pommier peut également augmenter la tolérance à des maladies comme le feu-bactérien causé par l'agent pathogène *Erwinia amylovora* (Jensen *et al.*, 2012), ou la pourriture racinaire induite par *Phytophthora cactorum* (Carisse & Khanizadeh, 2006).

Un autre intérêt majeur du greffage chez le pommier est la conservation des cultivars. En effet, du fait de son auto-incompatibilité pollinique, il n'est pas possible de multiplier une variété donnée à partir de pépins. Le greffage permet la conservation des variétés par la création de clones. Dans le monde, la quasi-totalité des pommiers cultivés sont issus du greffage (Fallahi *et al.*, 2002).

C'est cette caractéristique de multiplication clonale, comme source majoritaire de maintien de l'identité génétique, qui nous intéresse pour la suite de ce manuscrit.

3. La génétique chez les plantes

3.1. Le génome des plantes

Tous les organismes vivant (hors virus) connus à ce jours (Hiyoshi *et al.*, 2011), possèdent de l'acide désoxyribonucléique (ou ADN) au sein de leurs cellules. La molécule d'ADN est constituée de deux brins formant une double hélice (Figure 3.1-1 et Figure 3.1-2A). Chaque brin est composé d'une succession de nucléotides: l'adénine (A), la cytosine (C), la guanine (G) et la thymine (T). Les nucléotides s'apparient de façon spécifique deux à deux entre les deux brins de l'ADN pour former une double hélice (Figure 3.1-2A, Alberts (2003)). Ainsi une adénine s'apparie toujours avec une thymine et une cytosine s'apparie toujours avec une guanine.

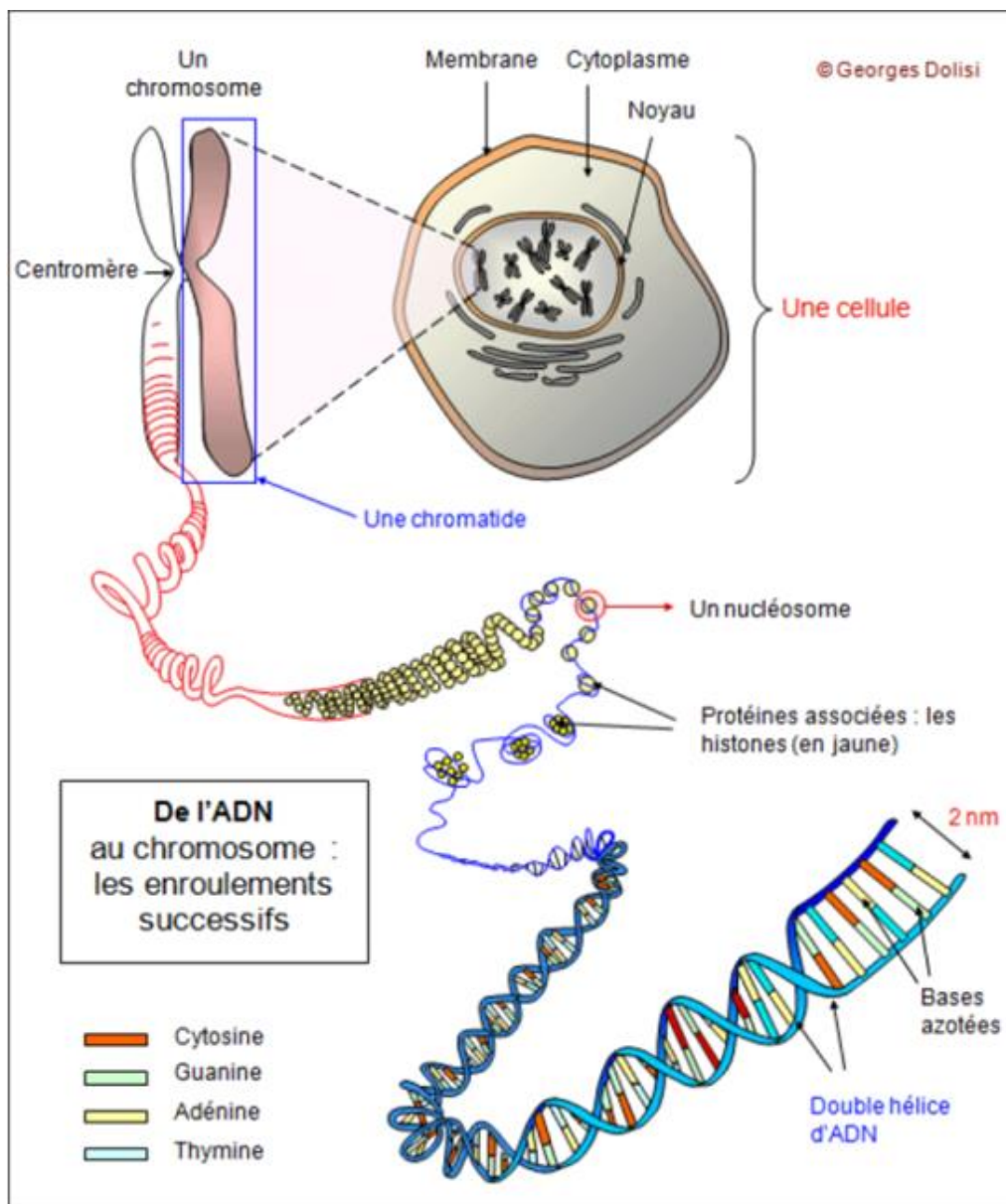


Figure 3.1-1 : structure et repliement de l'ADN (illustration par Georges Dolisi)

Une séquence d'ADN est donc définie par un enchaînement de nucléotides. La taille d'une séquence est alors exprimée en paires de bases (bp). La composition et l'ordre des nucléotides composant la séquence ADN d'un organisme est déterminée par des techniques de séquençage. L'analyse et le traitement des données de séquençage permet de construire une séquence qui servira de référence (génome de référence, (Ballouz *et al.*, 2019)) pour d'autres études sur la même espèce ou sur des espèces apparentées.

Les plantes sont des Eucaryotes (Evers *et al.*, 2013), ce qui signifie que l'ADN est enfermé dans le noyau de la cellule (Figure 3.1-1). Au sein du noyau, les molécules d'ADN interagissent avec des protéines (histones) qui permettent notamment leurs repliements afin de compacter l'ADN. Ces ensembles compactés forment ce que l'on appelle les chromosomes (Figure 3.1-1).

Le génome d'un organisme comporte toute l'information génétique nécessaire au développement, à la vie et à la reproduction de l'organisme. Cette information est codée dans la séquence d'ADN et est identifiée par des séquences nommées gènes (Figure 3.1-2A). Chaque gène permet, via la transcription puis la traduction, la synthèse d'une ou plusieurs protéines nécessaires au développement d'une cellule ou d'un organisme.

Chez les eucaryotes plusieurs parties ont été définies sur un gène (Figure 3.1-2A). Tout d'abord un site d'initiation de la transcription (ou TSS), un ou plusieurs exons, de zéro à plusieurs introns et un site de terminaison de la transcription (ou TTS). En amont et en aval du gène (aussi nommé région 5' et 3' respectivement), sont définies des zones nommées région promotrice ou terminatrice respectivement, sur lesquelles sont présentes des séquences qui vont permettre la régulation de la transcription du gène, notamment par la fixation de facteurs de transcription.

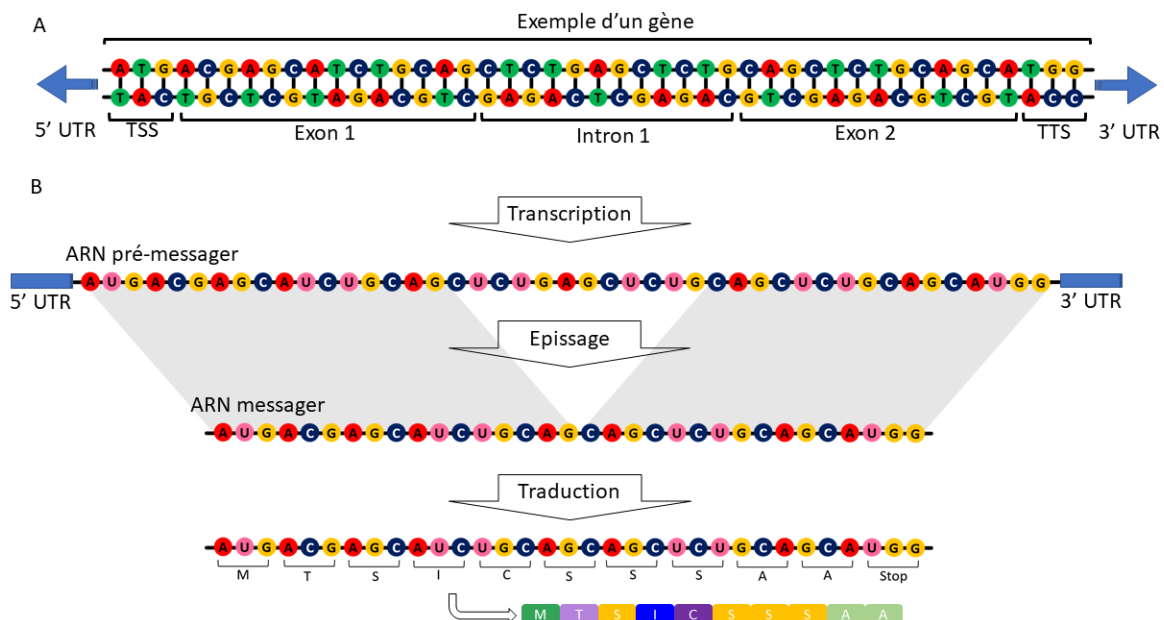


Figure 3.1-2: représentation de l'ADN et des étapes de transcription et traduction. (A.) Représentation d'un gène. UTR : région non traduite. (B.) représentation des étapes d'expression d'un gène.

3.1.1. Le génome du pommier

L'ADN (ou génome) du pommier a été séquencé pour la première fois en 2010 (Velasco *et al.*, 2010). Depuis 2010, plusieurs équipes dans le monde ont reséquencé le génome du pommier pour améliorer la qualité du génome de référence (Li *et al.*, 2016; Daccord *et al.*, 2017; Zhang *et al.*, 2019). Entre ces différentes versions du génome, la taille totale du génome varie. Pour la suite de ce manuscrit, nous considérerons le séquençage du génome du pommier 'GDDH13' (Golden Delicious double haploïde, clone numéro 13) réalisé par Daccord *et al.* (2017) qui estimaient la taille du génome à environ 650 mégabases (MB). La comparaison entre le génome de GDDH13 séquencé par Daccord et al 2017 et d'autres génomes précédemment séquencés est disponible dans le Tableau 3.1-1.

Tableau 3.1-1: Table issus de la thèse de Daccord et al . 2018 présentant la comparaison de l'assemblage du génome de GDDH13 avec deux assemblage réalisé précédemment sur pommier.

	This study	Velasco et al. (2010)	Li et al. (2016)
Total Number of Bases (Mb)	643.2	603.9	632.4
N50 (Kb)	5,558	16	112
Annotated protein coding genes	42,140	63,541	53,922
Transposable elements proportion (%)	57.3	42.4	NA
Pearson correlation coefficient with genetic map	0.90	0.67	NA
Complete BUSCOs	94.9%	86.7%	51.5%
Fragmented BUSCOs	2.6%	5.6%	18.8%
Missing BUSCOs	2.5%	7.7%	29.7%

Contrairement aux autres sous-familles des Rosacées, comportant un nombre chromosomique de base entre 7 et 9, les Méloïdés, dont fait partie le pommier, possèdent un nombre chromosomique de base de 17. Le pommier étant par ailleurs diploïde, son nombre de chromosomes est donc doublé et ces chromosomes sont appariés deux à deux : le pommier possède 17 paires de chromosomes. Cette augmentation du nombre de chromosomes par rapport aux autres sous-familles des Rosacées est due à une duplication relativement récente du génome (> 50 millions d'années, (Velasco *et al.*, 2010). Cette duplication a pour conséquence une forte homologie entre chromosomes/zones dupliqués dans le génome, comme montré par Daccord et al. (2017), dans le diagramme de synténie présenté en Figure 3.1-3.

Au sein du génome du pommier 42.120 séquences ont été annotées comme étant des gènes codant pour des protéines (Daccord *et al.*, 2017).

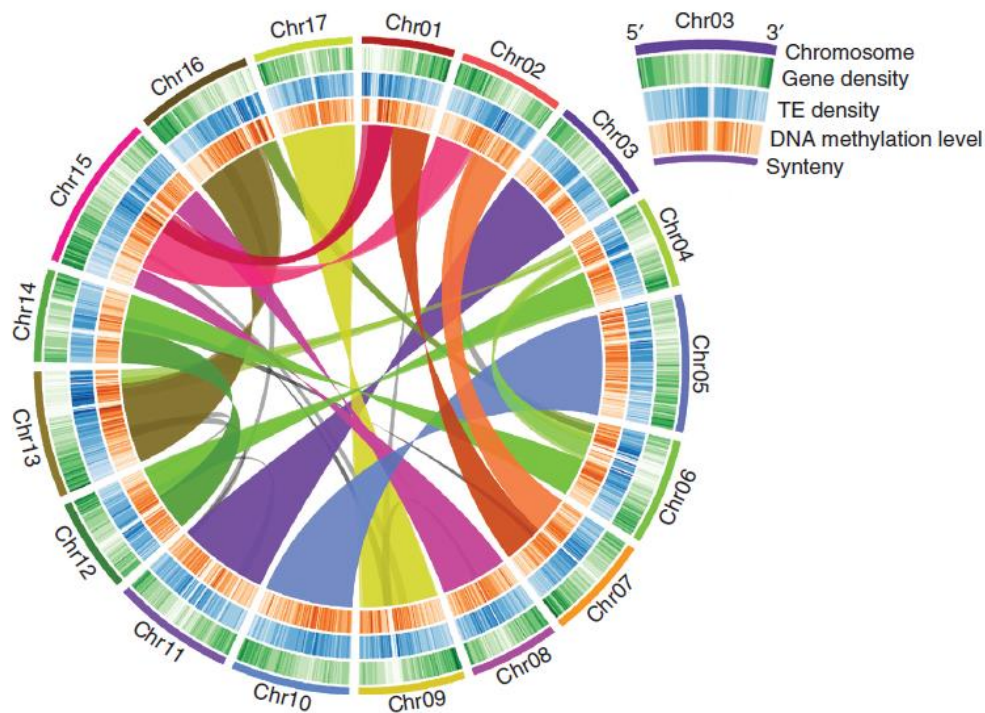


Figure 3.1-3 : diagramme de syntenie du génome du pommier proposé par Daccord et al 2017 : « Figure 2 Synteny and distribution of genomic and epigenomic features of the apple genome. The rings indicate (from outside to inside, as indicated in the inset) chromosomes (Chr), heat maps representing gene density (green), TE density (blue) and DNA methylation levels (orange). A map connecting homologous regions of the apple genome is shown inside the figure. The colored lines link collinearity blocks that represent syntenic regions that were identified by SynMap »

Le pommier étant diploïde, chaque gène possède deux allèles, un sur chaque chromosome apparié. Un allèle est une version d'un gène. Pour un organisme homozygote tous les allèles d'un gène seront identiques. Pour un organisme hétérozygote, tel que le pommier, tout ou partie des gènes ont des allèles différents. Chez le pommier par exemple, la couleur de l'épiderme des fruits rouges est codée par plusieurs gènes (Honda *et al.*, 2002), dont un régulant l'accumulation d'anthocyanine (Janick *et al.*, 1996). En fonction de l'allèle exprimé (transcription et traduction), dans l'épiderme des pommes, la couleur peut ainsi être modifiée (White & Lespinasse, 1986).

Des travaux ont été réalisés à l'INRAE d'Angers (IRHS) afin d'obtenir des pommiers présentant un génome homozygote. Ils ont permis l'obtention du pommier homozygote GDDH13 (Lespinasse *et al.*, 1999). Ce pommier GDDH13 a la particularité de ne pas être auto-incompatible. Il est donc possible d'en réaliser des autofécondations.

a) Les versions du génome de GDDH13

Lors du séquençage du génome du pommier différentes versions d'assemblage ont successivement été mises à disposition. Au cours de sa thèse, Nicolas Daccord (2018) a ainsi produit deux versions du génome de GDDH13 : la version V1.0 et la version V1.1 (Figure 3.1-4). La version V1.1 a été obtenue en réalisant un réassemblage par itérations successives (Voir la thèse de Nicolas Daccord 2018 pour plus de détails). Ceci permet d'améliorer l'assemblage, notamment pour les séquences répétées difficiles à assembler.

	GDDH13 v1	GDDH13 v1.1
base number	625275728	624851326
N number	85686323	84710065
coding genes	42140	45116
snoRNA	4	410
rRNA	123	4369
tRNA	510	654
miRNA	27	141
snRNA	0	181
other ncRNA	0	54
undefined ncRNA	1301	1816
total genes	44105	52741
complete BUSCO	94.9%	96.8%
fragmented BUSCO	2.6%	1.3%
missing BUSCO	2.5%	1.9%
perfect CDS (800 curated genes)	73.6%	75.1%
RNAseq evidence	93.2%	89%
PFAM signature	76.1%	74.8%

Figure 3.1-4 : Comparaison du réassemblage et de la ré-annotation des version V1.0 et V1.1 du génome de GDDH13. Source : <https://iris.angers.inra.fr/gddh13/the-apple-genome-downloads.html>

3.2. Transcription de l'ADN

Pour être utilisée par la cellule, l'information génétique doit tout d'abord être transcrite. La transcription est permise par des polymérase à ARN (RNA en anglais, Tjian, (1996)). Ce sont des enzymes spécifiques qui vont recopier un gène ou une séquence d'ADN pour former un brin d'ARN (Figure 3.1-2B). L'ARN est une molécule similaire à l'ADN, mais se composant d'un brin unique et pour laquelle la thymine est remplacée par l'uracile (U). Il existe différents types d'ARN définis par leur fonction dans la cellule.

3.2.1. Les ARNs messagers ou ARNs codants

Les l'ARNs messagers (mRNA) sont transcrits à partir des séquences des gènes, par les polymérase qui vont former dans un premier temps un ARN pré-messager (Figure 3.1-2B). Cet ARN pré-messager va subir un épissage (Darneli, 1978; Black, 2003) qui permet de supprimer les introns, d'ajouter une coiffe sur la régions en 5' et une queue poly-A en 3' permettant la stabilité de l'ARNm. L'épissage peut également modifier le nombre d'exons retenu dans le mRNA final. Cet épissage alternatif permet à la cellule de former plusieurs mRNA différents à partir d'une seule séquence d'ADN initiale.

3.2.2. Les ARNs non-codants

Les ARNs non-codants sont impliqués dans divers mécanismes. On retrouve principalement les ARNs de transferts (tRNA) et les ARNs ribosomiaux (rRNA) qui permettent tous deux la traduction des mRNAs en protéines (Sharp *et al.*, 1985). Les microARNs (miRNA, (Bartel, 2004)), qui regroupent les ARNs interférant (iRNA) et les petits ARNs interférant (siRNA), sont impliqués dans la régulation post-transcriptionnelle des gènes. Ou encore les petits ARNs nucléaire

(snoRNAs, Maden & Hughes, (1997)) qui sont impliqués dans la maturation des tRNAs et rRNAs. Il existe de nombreuses autres classes de petits ARNs impliqués dans des régulations spécifiques (Xie *et al.*, 2004; Mosher *et al.*, 2010).

3.3. Traduction de l'ARNm

Suite à la transcription et à la maturation du mRNA, cette molécule est transportée à l'extérieur du noyau de la cellule, dans le cytoplasme, pour y être traduite (Figure 3.1-2B et Figure 3.3-1A). La traduction est assurée par des ribosomes (Wilson & Cate, 2012). Ce sont des complexes formés par les rRNAs et des protéines. Ils permettent de traduire une séquence d'ARNm en protéine. Lors de la traduction, à chaque triplet de nucléotides (ou codon) va correspondre un acide aminé particulier (Figure 3.3-1B), la correspondance de chaque codon en acide aminé forme le code génétique (Crick, 1990). Le ribosome va donc lire l'ARN messager et produire une chaîne d'acide aminé qui formera une protéine qui aura une fonction particulière pour la cellule. L'ensemble des protéines produites permettent entre autres le développement et la multiplication des cellules, leur défense, mais aussi la communication entre cellules (Harvey *et al.*, 2007).

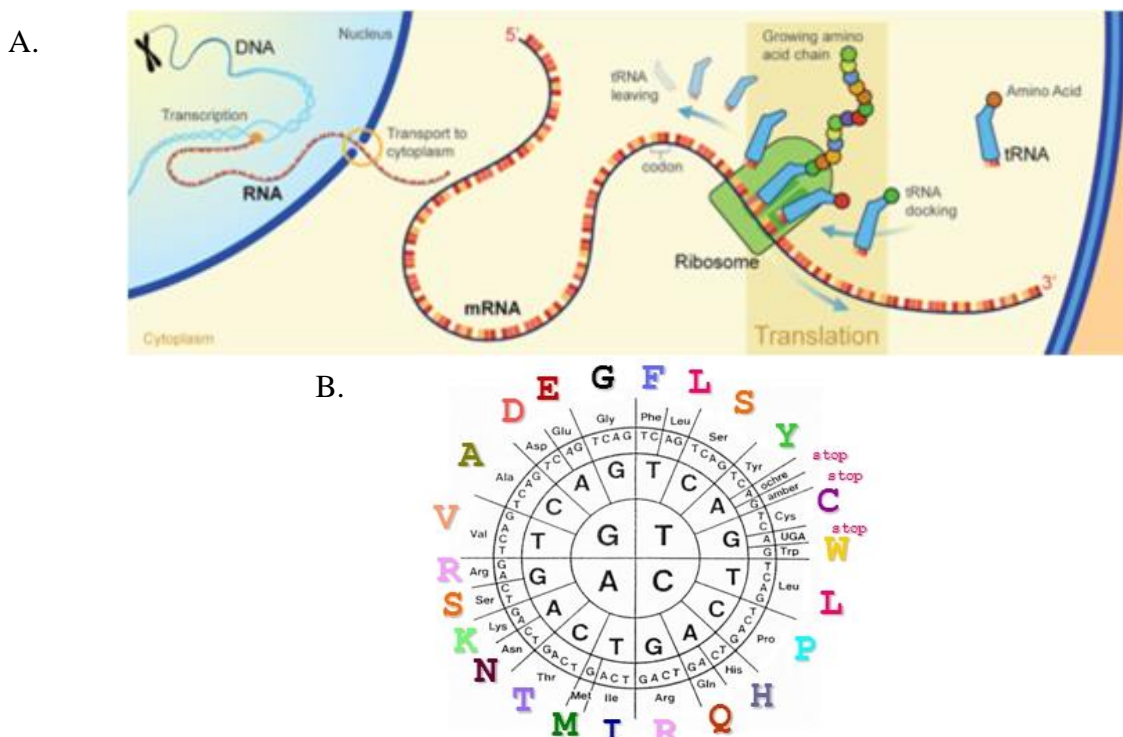


Figure 3.3-1: La traduction : (A.) Schéma de la traduction des ARNm par les ribosomes dans le cytosol. Source : <https://www.ck12.org>. (B.) Code génétique. Pour trouver l'acide aminé correspondant à un codon de nucléotide, il faut partir du centre et aller vers l'extérieur. Source: <http://education.expasy.org>.

3.4. Régulation de l'expression des gènes

Une cellule ne produit pas en permanence les protéines codées par l'ensemble de son génome, mais seulement la fraction nécessaire à sa vie et à son développement à un instant donné. Pour les organismes complexes, les cellules peuvent être regroupées en sous-ensemble en fonction de l'organe (feuille, racine, tige) ou de leur tissus (tissus conducteur, tissus photosynthétique, tissus assimilateur, etc...). Ces cellules transcrivent ainsi un certain nombre de gènes en fonction de leur spécificité. Afin de réguler l'expression des gènes un certain nombre de mécanismes entrent en jeu à différents niveaux.

3.4.1. Régulation génétique

Il existe des protéines qui ont un rôle d'activation ou de répression de la transcription et qui sont appelées des facteurs de transcription. Ces protéines vont reconnaître des séquences particulières de l'ADN, situées notamment dans les régions promotrices ou terminatrices des gènes. La fixation de ces protéines va permettre le recrutement des ARN polymérases dans le cas d'activateurs de la transcription, ou empêcher la fixation des ARN polymérases dans le cas de répresseurs de la transcription. L'une des familles de facteurs de transcription les plus connues est la famille des MYB, qui est largement répandue notamment chez les plantes (Martin & Paz-Ares, 1997). Par exemple chez le pommier, l'implication de facteurs de transcription de type MYB a été décrite dans la régulation de l'accumulation d'anthocyanes (pigment rouge) dans les fruits (Takos *et al.*, 2006; Espley *et al.*, 2007; Ban *et al.*, 2007).

3.4.2. Régulation post-transcriptionnelle

Un second niveau de régulation de l'expression des gènes peut intervenir après la transcription. Une partie des mRNAs produits par la cellule va subir un processus de régulation pour être détruit avant la traduction et ainsi empêcher la formation de la protéine associée. Cette régulation est notamment rendue possible par l'action des miRNAs.

Les miRNAs sont des petits ARNs de 20 à 24 nucléotides impliqués dans la régulation post-transcriptionnelle (Bartel, 2004). Ils ont été identifiés en 1993 chez *Caenorhabditis elegans* (Lee *et al.*, 1993). Les miRNAs font partie d'un complexe ribonucléoprotéique de répression nommé RISC (« RNA-induced silencing complex »). Dans ce complexe sont notamment retrouvées les protéines DICER-like (impliquées dans la maturation des miRNA, Liu *et al.*, (2009)) et Argonaute (permettant le clivage des mRNA, Vaucheret, (2008)). Une fois le complexe formé le miRNA peut se fixer sur son ARNm cible pour induire sa dégradation via le complexe RISC (Bartel, 2004).

En plus des différents mécanismes de régulation génique présentés plus haut, il existe également des mécanismes de régulation qui ne sont pas codés par la séquence d'ADN, ils sont appelés mécanismes de régulation épigénétiques.

4. Régulation épigénétique chez les plantes

4.1. Définition de l'épigénétique

L'épigénétique se définit comme l'ensemble des modifications héritables via la méiose et la mitoses, pouvant modifier l'expression des gènes, sans être liées à une modification de la séquence d'ADN (Russo *et al.*, 1996). Ces modifications sont héritables par la méiose (division cellulaire nécessaire à la reproduction sexuée) ou par la mitose (multiplication cellulaire). Les deux types de modifications épigénétiques majeures sont les modifications d'histones et la méthylation de l'ADN. Les modifications épigénétiques ont un rôle dans la régulation de l'expression des gènes et vont donner une plasticité à l'expression de ces gènes (Pikaard & Scheid, 2014). Cette plasticité permet une adaptation aux contraintes environnementales auxquelles sont soumis les organismes (Richards *et al.*, 2010; Mirouze & Paszkowski, 2011; Casadesús & Low, 2013).

4.2. Les histones

4.2.1. Définition

Les histones sont des complexes de protéines autour desquelles l'ADN (Figure 4.2-1A) s'enroule pour former des nucléosomes (Figure 3.1-1, Peterson & Laniel, (2004)). Les histones permettent de compacter l'ADN au sein du noyau, mais également d'organiser cette compaction de telle sorte que l'ADN reste accessible pour la transcription. La chromatine accessible à la transcription est qualifiée d'euchromatine (revue dans Peterson & Laniel, (2004)). En augmentant la compaction de l'ADN les histones peuvent rendre l'ADN moins accessible à la transcription, on parle alors d'hétérochromatine (Shu *et al.*, 2012).

Pour permettre le passage d'un état décondensé à un état condensé et inversement, les histones sont la cible de nombreuses modifications (acétylation, méthylation, ubiquitination, etc...) qui permettent de réguler leur activité ou l'accessibilité des enzymes à l'ADN (revue dans Peterson & Laniel, (2004), Lister & Ecker, (2009)). Ces modifications ont principalement lieu sur la queue N-terminale des histones (Figure 4.2-1A), qui est une partie de la séquence protéique de l'histone qui reste libre. Au sein d'un nucléosome on trouve 8 protéines de type histones, chacune de ces histones ayant des fonctions particulières (Talbert & Henikoff, 2010). Les modifications ayant lieu sur les queues d'histones vont cibler spécifiquement les différents types d'histone pour avoir une action particulière.

4.2.2. Modification des histones

La queue des histones peut donc être la cible de modifications qui consistent en l'ajout ou la suppression de groupements biochimiques (Figure 4.2-1B).

Il existe différents types de modifications chimiques dont l'acétylation et la méthylation (Loidl, 2004).

a) L'acétylation de la queue histone

L'acétylation est une marque qui peut être ajoutée ou retirée via des histones acétylases ou désacétylases. L'acétylation a lieu sur une lysine spécifique présente dans la séquence protéique de la queue de l'histone. Une acétylation de la queue d'histone est généralement associée à l'euchromatine, transcriptionnellement active (Berger, 2007). Alors que la désacétylation des

histones est associée à l'hétérochromatine (Berger, 2007). Cependant des études ont montré que l'acétylation des histones est dynamique en fonction du développement d'une plante, et que l'acétylation peut moduler la transcription des gènes positivement ou négativement (revue par Chen & Tian, (2007)).

b) La méthylation de la queue histone

Tout comme l'acétylation, des groupements méthyl peuvent être déposés sur les lysines composant la queue des histones. A l'inverse de l'acétylation, la méthylation des histones est généralement associée à la répression de la transcription des gènes, mais là encore cet effet est très modulable (Berger, 2007). En effet la tri-méthylation de la lysine située en quatrième position sur la queue d'une histone nommée H3 a un effet activateur sur la transcription des gènes, alors que la tri-méthylation de la lysine situé en neuvième position de cette même histone est associée à des régions d'hétérochromatine ayant un faible état transcriptionnel (Berger, 2007).

c) Acétylation et méthylation des histones chez le pommier

Chez le pommier, 57 histones acétyltransférases et 26 histones désacétylases ont récemment été identifiées (Fan *et al.*, 2018) sur le génome de référence 'GDDH13' (Daccord *et al.*, 2017). Pour la méthylation de la queue des histones, 71 histones méthyltransférases et 44 histones déméthylases ont été identifiées par Fan *et al.*, (2018) sur le génome de référence 'GDDH13' (Daccord *et al.*, 2017). Les modifications d'histones chez les arbres, tel que le pommier, ne sont pas encore suffisamment étudiées. Des études ultérieures seront nécessaires pour vérifier le rôle chez les arbres.

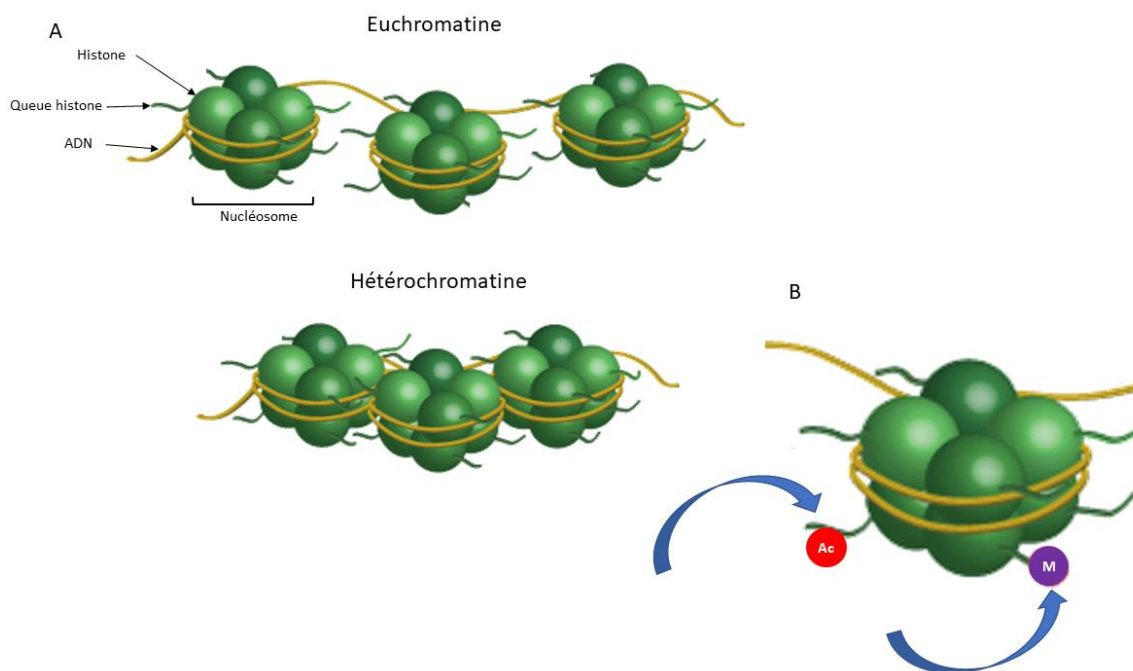


Figure 4.2-1: Illustration d'histones. (A) forme condensée (hétérochromatine) et décondensée (euchromatine) de la chromatine. (B) exemple de modification sur une queue d'une histone. Adaptée de www.epigentek.com.

Quelques soient leurs effets activateurs ou inhibiteurs sur la transcription des gènes, il a également été montré que l'activité des histones est intimement liée à la méthylation de l'ADN, notamment via le recrutement par les histones d'enzymes impliquées dans la méthylation de l'ADN (Dobosy & Selker, 2001; Jackson *et al.*, 2002; Malagnac *et al.*, 2002).

4.3. La méthylation de l'ADN

4.3.1. Définition

La méthylation de l'ADN consiste en l'addition d'un groupement méthyle (CH_3) sur une base azotée de l'ADN (Figure 4.3-1). Chez les plantes les adénines et les cytosines peuvent être méthylées (Finnegan *et al.*, 1998; Ashapkin *et al.*, 2002; Chan *et al.*, 2005; Ratel *et al.*, 2006), cependant la méthylation est prédominante sur les cytosines (Wyatt, 1951).

Pour la suite de ce manuscrit je parlerais exclusivement de la méthylation de l'ADN sur les cytosines. On notera cependant que la méthylation des adénines est associée à la régulation de la transcription des gènes mitochondriaux (Ashapkin *et al.*, 2002).

La méthylation des cytosines peut avoir lieu dans trois contextes de cytosines différents, définit par la ou les bases succédant à la cytosine : les contextes CG, CHG et CHH, où H peut être un nucléotide A, T ou C (Vanyushin & Ashapkin, 2006). Le niveau de méthylation moyen de chaque contexte est variable d'une espèce à une autre. Globalement le contexte CG est souvent le plus méthylé, suivi du contexte CHG et enfin du contexte CHH, qui est très peu méthylé, comme montré chez *Arabidopsis thaliana* (respectivement un taux de 24%, 6.7% et 1.7% de méthylation, *A. thaliana*, Cokus *et al.*, (2008)), *Glycine max* (respectivement un taux de ~50%, ~39% et ~5,3% de méthylation, *G. max*, Schmitz *et al.*, (2013)).

a) Chez le pommier

Lors du séquençage du génome du pommier en 2017, Daccord *et al.*, ont également effectué un séquençage du méthylome. Ils ont ainsi pu mettre en évidence un profil de méthylation similaire à *G. max* (Schmitz *et al.*, 2013). Ainsi en contexte CG un taux de méthylation de 49% a été identifié, 39% en contexte CHG et 12% en contexte CHH (Daccord *et al.*, 2017).

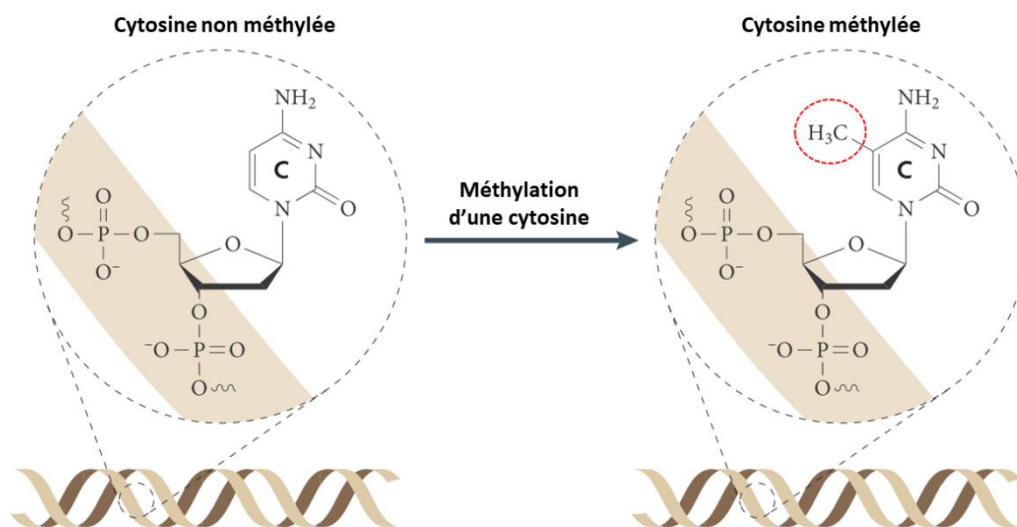


Figure 4.3-1: représentation de la méthylation d'une cytosine à l'échelle moléculaire, par ajout d'un groupement CH_3 (adaptée de (Law & Jacobsen, 2010))

4.3.2. Rôle de la méthylation de l'ADN

Ces modifications ne sont pas réparties de façon aléatoire dans le génome (Michalowsky & Jones, 1989; Lister *et al.*, 2008). Tout au long de la vie de la plante, la méthylation est très dynamique. La méthylation de l'ADN de la plante est notamment associée à la réponse aux un stress biotiques ou abiotiques (Kinoshita & Seki, 2014; Viggiano & de Pinto, 2017; Thiebaut *et al.*, 2019). La perception d'un stress par la plante entraîne une plasticité phénotypique et transcriptomique lui permettant de s'adapter. Cette perception s'accompagne également d'une modification de l'épigénome et notamment de la méthylation de l'ADN (Downen *et al.*, 2012a; Iwasaki *et al.*, 2014). En 2017, Seymour & Becker s'interrogeaient sur la relation de cause à effet : est-ce la plasticité phénotypique qui entraîne une modification épigénétique ou bien la perception d'un stress engendre-t-elle des modifications directes de l'épigénome se traduisant ensuite par une plasticité transcriptomique et phénotypique ?

La méthylation joue également un rôle dans la stabilité du génome via la répression de la transcription des éléments transposable (TE) (Kawashima & Berger, 2014).

a) Méthylation de l'ADN et éléments transposables

Un TE est une séquence capable de s'intégrer dans une séquence d'ADN par transposition (Feschotte *et al.*, 2002). Ce sont des séquences d'ADN répétées impliquées dans l'augmentation de la taille des génomes chez les eucaryotes (Kidwell, 2002). On observe d'ailleurs que leur proportion dans les génomes augmente avec la taille du génome. Par exemple pour l'espèce *Utricularia gibba*, dont le génome a une taille de 82Mb, les TEs représentent 3% du génome (Ibarra-Laclette *et al.*, 2013), alors que pour le maïs (*Zea mays*), dont la taille du génome est de 2,2 Gb, les TEs représentent 85% du génome (Schnable *et al.*, 2009).

La mobilité des TEs génère des mutations et est une source importante d'évolution des génomes et des gènes (Bennetzen, 2005), et par conséquent du phénotype. Chez le maïs, il a été montré que l'insertion d'un transposon à proximité d'un locus impliqué dans la synthèse d'anthocyanine altère la production de pigments dans le grain. Cette insertion provoque une modification de la couleur des grains de maïs qui peuvent alors être noir, jaune, ou former une mosaïque (noir et jaune) (revu dans Wessler, (1988)). Un phénotype associé à la couleur du grain, causé par l'insertion d'un TEs dans un gène a également été décrit chez le Shorgo (*Sorghum bicolor*, Chopra *et al.*, (1999)). Chez le riz, on retrouve également que l'expression d'un gène impliqué dans le nanisme et la mise en place des feuilles (*Oryza sativa*) est régulé par des éléments transposable nommés MITEs (« Miniature Inverted repeat TEs », Wei *et al.*, (2014)).

Cependant une activité (mobilisation) trop intense de ces séquences d'ADN est néfaste pour la survie des cellules et de l'organisme. Ceci a notamment été montré chez *A. thaliana* où la transposition d'un TE nommé « Evadé » est responsable du développement de phénotypes aberrants (Mirouze *et al.*, 2009). Cet effet délétère de la transposition de TEs semble aussi être corroboré par l'observation de la répartition des TEs dans le génome. En effet les TEs sont majoritairement retrouvés dans des régions intergéniques chez *A. thaliana* (Wright *et al.*, 2003). Cependant des nuances existent, en effet les éléments MITEs du riz sont quant à eux fréquemment retrouvés à proximité des gènes (Bureau & Wessler, 1994; Zhang *et al.*, 2000; Yang *et al.*, 2001). Des études suggèrent également que le choix de la multiplication sexuée ou asexuée pour un organisme est relié au taux de transposition des TEs (Arkhipova & Meselson, 2005).

L'activité des TEs doit donc être régulée finement par la cellule. Il existe pour cela des mécanismes de répression de la transcription des TEs, telle que la méthylation de l'ADN (Feschotte *et al.*, 2002; Goll & Bestor, 2005; Henderson & Jacobsen, 2007; Suzuki & Bird, 2008; Hauser *et al.*, 2011; Schmitz *et al.*, 2012). Les éléments transposables sont la cible de petit ARNs interférant (siRNA) permettant la méthylation de l'ADN correspondant (Almeida & Allshire, 2005) mais également la régulation post-transcriptionnelle de l'ARN correspondant.

On notera enfin que la méthylation des TEs insérés à proximités d'un gène joue un rôle dans la modulation de la transcription de ces gènes (Slotkin & Martienssen, 2007).

Méthylation des éléments transposable chez le pommier

Nous avons vu précédemment que la taille d'un génome de plante est corrélée à l'abondance en TEs qu'il contient. Ainsi chez le pommier, qui a un génome de taille moyenne (650Mb), les TEs représentent 47,3% du génome (Daccord *et al.*, 2017). Xu *et al.* 2018 ont confirmé que la méthylation de l'ADN assure la stabilité du génome du pommier.

On notera également que, chez le pommier, l'insertion d'un TE en amont du gène codant le facteur de transcription MdMYB1, permet de réguler son niveau de transcription et donc celui de son gène cible, qui est impliqué dans la pigmentation rouge de l'épiderme des pommes (Zhang *et al.*, 2019).

Méthylation de l'ADN et régulation de la transcription des gènes

La méthylation de l'ADN permet la régulation de la transcription de certains gènes (Michalowsky & Jones, 1989; Inamdar *et al.*, 1991), notamment lorsqu'elle touche les régions promotrices (Bird, 1984; Cedar, 1988; Di Croce *et al.*, 2002; Baylin, 2005). Cependant l'effet de la méthylation de l'ADN sur la régulation de l'expression des gènes est variable en fonction du contexte de méthylation ou de la localisation de la méthylation. Comme on peut le constater sur le modèle décrit par Niederhuth & Schmitz, (2016), Figure 4.3-2. Niederhuth & Schmitz, (2016) décrivent ainsi que la méthylation du corps du gène (Figure 4.3-2A) dans les trois contextes à un effet inhibiteur sur la transcription de ce gène. Ils décrivent ensuite les effets de deux localisations distinctes de la méthylation, en contexte CG uniquement. Tout d'abord une forte méthylation dans le corps du gène (Figure 4.3-2B) aurait un effet plutôt activateur sur la transcription, alors qu'une forte méthylation au niveau du site d'initiation de la transcription aurait plutôt un effet inhibiteur sur la transcription du gène (Figure 4.3-2C).

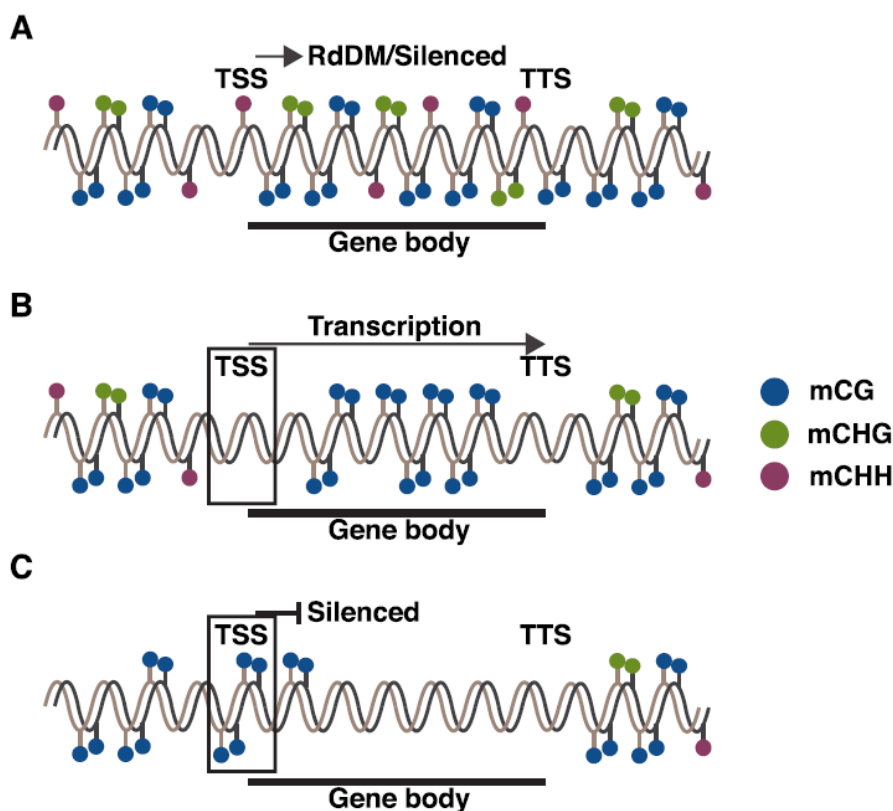


Figure 4.3-2: Modèle de régulation de l'expression de gènes d'*A.thaliana* via la méthylation de l'ADN. Figure 1 de la publication de Niederhuth CE et Schmitz RJ. (2016)

Cette régulation peut s'expliquer par le fait que la méthylation permet de bloquer la fixation de facteurs de transcriptions ou des enzymes responsables de la transcription (Adams, 1990). Ainsi la méthylation d'une séquence CG adjacente à une séquence TATA (« TATA box ») va empêcher la fixation des RNA polymérase sur la séquence TATA, limitant ainsi la transcription des gènes (Adams, 1990).

Downen *et al.* (2012) ont montré chez *A. thaliana* qu'une modification de l'épigénome corrélée à une modification de la transcription de gènes de défense est observable en cas de stress biotique. Ces modifications peuvent être durables et permettraient à la plante de répondre plus rapidement à un stress ultérieur (Avramova, 2015; Fortes & Gallusci, 2017).

Méthylation et régulation de l'expression des gènes chez le pommier

Chez le pommier la méthylation de l'ADN est prédominante sur les TEs et est plus faible sur les gènes. En effet 64,5% des gènes sont pauvres en méthylation (c'est-à-dire un taux de méthylation inférieur à 0,5%) et seul 10% des gènes sont fortement méthylés (c'est-à-dire un taux de méthylation supérieur à 2,5%) (Daccord *et al.*, 2017).

Daccord *et al.* (2017) ont également mis en évidence que les gènes fortement méthylés ont un niveau de transcription faible, alors qu'un faible taux de méthylation des gènes est associé à un niveau de transcription plus élevé, ce qui est également confirmé par Xu *et al.*, (2018). Dans leur publication, Daccord *et al.* (2017) présentent notamment deux clones d'un 'Golden Delicious' haploïde doublé (GDDH13 et GDDH18). Ces deux arbres présentent un génome similaire, cependant on peut observer une différence phénotypique au niveau de la taille des fruits. L'analyse

du méthylome de ces fruits a révélé de larges différences de méthylation au niveau de gènes impliqués dans le développement et la maturation du fruit (Daccord *et al.*, 2017).

De plus, comme déjà évoqué précédemment, il a été montré que le niveau de méthylation du promoteur du facteur de transcription MYB1 a un effet sur l'accumulation d'anthocyanine dans la peau des pommes, via la modification de l'expression des gènes cibles de MYB1 (Ma *et al.*, 2018 ; Zhang *et al.*, 2019). Il a également été suggéré que la méthylation du promoteur du gène *IPT5b* chez le pommier est associée au caractère de nanisme induit par le porte greffe nommé M9. Cette méthylation produit une modification du profil d'expression de ce gène impliqué dans la voie de biosynthèse des cytokinines (Feng *et al.*, 2016).

La méthylation permet donc une régulation de la transcription des gènes. Cette régulation peut également intervenir en cas de stress biotiques ou abiotiques. Xu *et al.*, (2018), ont montré qu'un stress hydrique chez le pommier modifie la méthylation à l'échelle du génome. Ils ont notamment mis en évidence que le stress hydrique provoque une diminution de la méthylation dans la région promotrice des gènes différenciellement exprimés lors de ce stress, notamment le gène *MdOCP3* connu pour être impliqué dans la tolérance au stress hydrique (Xu *et al.*, 2018).

Gully *et al.*, (2019a), ont effectué chez *A. thaliana* et *M. domestica* des traitements de « priming », c'est-à-dire d'induction de la réponse à un stress chez une plante, sans qu'elle ait été confrontée à ce stress. Ce traitement de « priming » a permis aux plantes de mieux tolérer le stress ultérieurement. Ce traitement de « priming » est potentiellement associé à des modifications du niveau de méthylation de l'ADN.

4.3.3. Mécanisme de méthylation de l'ADN

Différents mécanismes existent afin d'établir et de maintenir la méthylation des cytosines chez les plantes, en fonction du contexte dans lequel elles se situent.

a) La maintenance de la méthylation

La méthylation est dynamique tout au long de la vie d'une plante (Schmitz *et al.*, 2012) grâce à des méthylases et des déméthylases qui vont ajouter ou enlever les marques de méthylation en fonction des besoins de la cellule. Les mécanismes de maintenance de la méthylation de l'ADN permettent une répression de la transcription des gènes et des éléments transposables tout au long de la vie des cellules.

La maintenance de la méthylation permet également la méthylation du brin d'ADN nouvellement synthétisé lors de la réplication de l'ADN (Figure 4.3-3).

Les gènes impliqués dans la maintenance de la méthylation sont différents en fonction du contexte de méthylation considéré, selon qu'il est de nature symétrique (CG et CHG) ou asymétrique (CHH) (revu par (Law & Jacobsen, 2010)).

Pour les contextes CG et CHG, qui sont symétriques (Figure 4.3-3), les protéines impliquées dans la restauration de la méthylation de l'ADN vont s'appuyer sur la méthylation existante sur la cytosine du brin complémentaire. Ainsi pour le contexte CG, des protéines VIM1 et 2 (« VARIANT IN METHYLATION ») vont reconnaître les cytosines non-méthylées du brin nouvellement synthétisé et devant être méthylées. Ces protéines vont alors recruter la protéine MET1 (« DNA

METHYLTRANSFERASE 1 ») qui va restaurer la méthylation sur le brin néosynthétisé (Kim *et al.*, 2014). De façon analogue pour le contexte symétrique CHG, des protéines SUVH4, 5 et 6 (« SUPPRESSOR OF VARIATION 3-9 HOMOLOGUE », Li *et al.*, (2018)) vont détecter les cytosines à méthyler sur le brin néosynthétisé et vont recruter la protéine CMT3 (« CHROMOMETHYLASE 3 ») pour restaurer la méthylation (Lindroth, 2001).

Enfin pour le contexte asymétrique CHH, la méthylation ne peut pas être restaurée sur le brin néosynthétisé en utilisant l'autre brin comme modèle, car il n'y a pas de cytosine correspondante sur laquelle recueillir l'information. La restauration de la méthylation pour ce contexte passe donc constamment par la voie de la méthylation *de novo*, guidée par des petits ARNs (mécanisme de la RdDM, « RNA directed DNA methylation »).

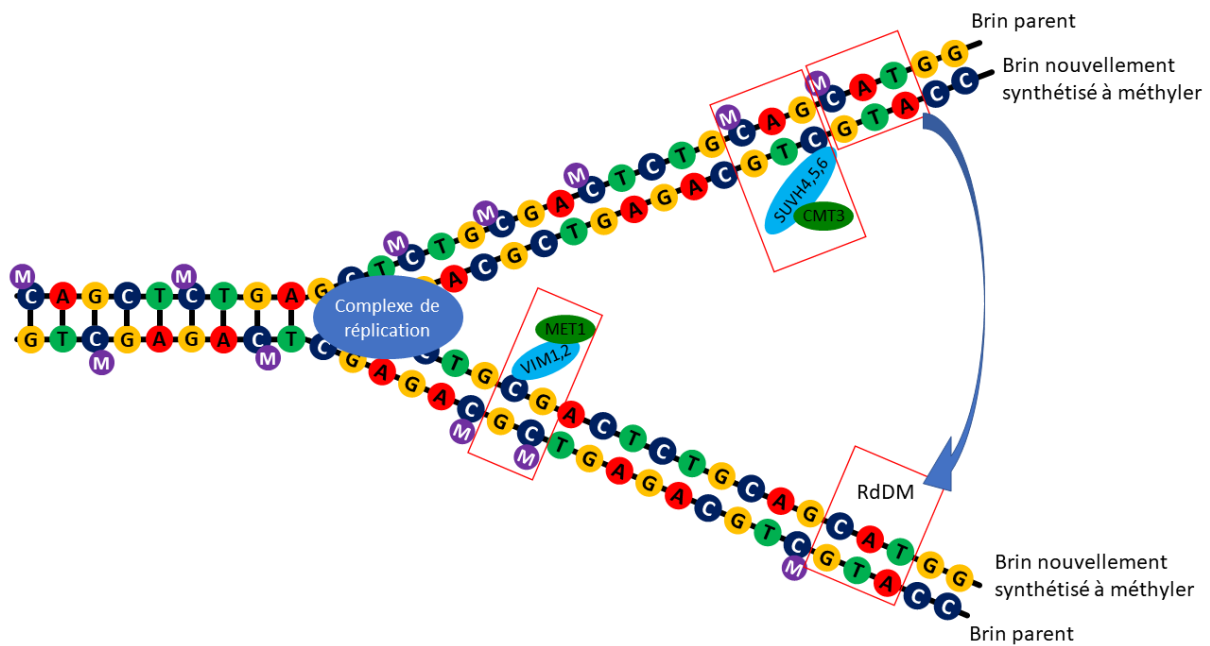


Figure 4.3-3: représentation des mécanismes de maintenance de la méthylation et de la voie de la RdDM, au niveau de la fourche de réplication. La méthylation des cytosines est représentée en violet.

b) La méthylation de novo

La méthylation *de novo* permet d'établir la méthylation d'une séquence ADN qui n'était pas méthylée précédemment. Elle a lieu tout au long du développement d'une plante. Elle permet notamment d'empêcher la transcription des TEs, assurant ainsi la stabilité et l'intégrité du génome (Wassenegger *et al.*, 1994).

La méthylation *de novo* repose essentiellement sur le mécanisme de la RdDM (Wassenegger *et al.*, 1994). Elle permet la méthylation des cytosines dans les trois contextes (CG, CHG et CHH). Elle nécessite le recrutement d'ARNs interférants de 24 nucléotides, complémentaires de la séquence à méthyler.

Le mécanisme de la RdDM comporte plusieurs étapes (revu par Law & Jacobsen, 2010) : il commence par la synthèse d'un ARN simple brin complémentaire de la région à méthyler par le complexe formant la polymérase IV (Pol IV, Figure 4.3-4, Onodera *et al.*, 2005). A partir de ces ARNs, un ARN double brin est formé par une RDR2 (Haag *et al.*, 2012), puis est modifié par une protéine de la famille des DICER (DCL3, Gascioli *et al.*, 2005), pour former un ARN interférant de 24 nucléotides (« 24nt siRNA »), qui sera pris en charge par des protéines Argonautes (*p.e.* AGO4, Vaucheret, (2008)). AGO4 est elle-même prise en charge par le complexe de la polymérase V (Pol V) qui va ensuite interagir avec DRM2 (« DOMAIN REARRANGED METHYLTRANSFERASE 2 », (Cao & Jacobsen, 2002)). C'est à la suite de ces interactions en cascade que DRM2 va transférer un groupement méthyle sur une cytosine de la région ciblée initialement.

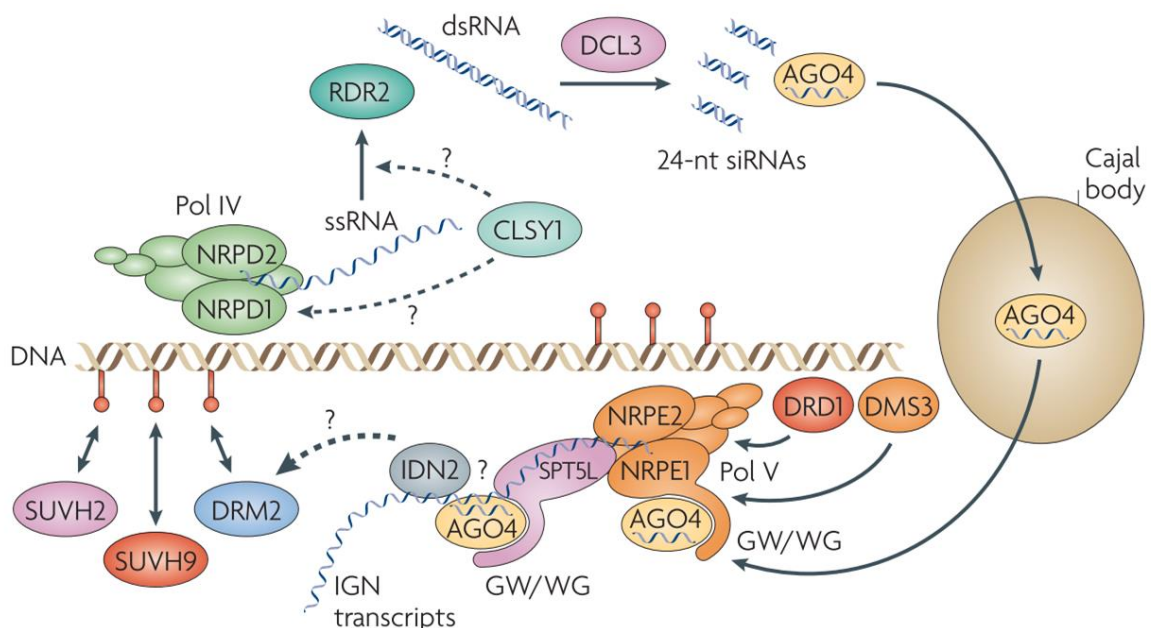


Figure 2 | **Model for RNA-directed DNA methylation.**

Figure 4.3-4: figure 2 de la revue de Law & Jacobsen, (2010) présentant les interactions moléculaires menant à la méthylation *de novo* de l'ADN via le mécanisme de la RdDM.

4.3.4. Méthylation de l'ADN et différences entre plantes pérennes et annuelles

Les plantes peuvent être divisées en deux catégories : les plantes annuelles qui ont un cycle de vie complet en un an et les plantes pérennes qui ont une période de vie plus longue. Les plantes pérennes doivent ainsi mettre en place des stratégies de développement différentes des plantes

annuelles. A partir de leur maturité elles doivent être capable de produire des fleurs et de les produire à un moment donnée précis de l'année, afin de pouvoir se croiser à l'intérieur de l'espèce. Ces différents processus sont notamment régulés par des modifications de la chromatine, comme montré chez *A.thaliana* par rapport à *A. alpina* (Wang *et al.*, 2009). Ces modifications de la chromatine associé au cycle de floraison chez les plantes pérennes ont également été décrite chez différentes espèces, notamment en comparant les bourgeons dormant des bourgeons débourrés (revu dans Ríos *et al.*, 2014).

Les plantes pérennes ont donc probablement développé des mécanismes de régulation épigénétiques différents des plantes annuelles afin de s'adapter à leur cycle de vie plus long et aux conditions climatiques récurrentes.

4.3.5. Méthylation de l'ADN et amélioration des plantes

Nous l'avons vu précédemment, le rôle de la méthylation de l'ADN sur l'expression des gènes est encore mal connu. La méthylation peut être activatrice ou inhibitrice de la transcription d'un gène, indépendamment ou non du contexte de méthylation. Ces différences ont notamment été formalisées par House & Lukens, 2019. Ils expliquent l'existence de trois type d'épiallèles. Les épiallèles purs qui sont des modifications épigénétiques sans liens avec l'ADN. Et les épiallèles « facilitated » et obligatoire, pour lesquels les modifications épigénétiques sont partiellement ou complètement dépendante de la séquence d'ADN sous-jacente. Il existe un regain d'attention pour ces deux dernières catégories de la part des obtenteurs. En effet, ces épiallèles sont considérés comme les plus stables et leur obtention est mieux comprise (House & Lukens, 2019). Par exemple Durand *et al.*, (2012) ont montré que l'incompatibilité entre différents variants *A.thaliana* est contrôlée par des modifications épigénétiques. Ils ont montré que l'insertion d'une séquence homologue à un gène de ménage fonctionnel conduit à la formation de siRNA par la RdDM, qui vont engendrer une méthylation du gène natif et une diminution de son expression. Ils ont également montré que cet épiallèle était stable à travers la méiose, sur plusieurs générations. Des travaux sur l'utilisation de l'épigénétique pour l'amélioration des plantes ont été réalisés chez différentes espèces d'intérêt agronomique tel que le colza (Hauben *et al.*, 2009) ou la tomate (Yang *et al.*, 2015)

Cependant les possibilités offertes aux obtenteurs par l'utilisation de l'épigénétique restent limitées au mécanisme épigénétiques maîtrisé actuellement, c'est-à-dire l'hyperméthylation d'une séquence ADN pouvant conduire à une diminution de l'expression des gènes. Une compréhension plus complète des mécanismes de la méthylation de l'ADN et de leurs effets devrait permettre à l'avenir une utilisation élargie des modifications épigénétiques pour l'obtention de nouvelles variétés (Xu *et al.*, 2018a).

5. Transmission de l'information épigénétique

L'une des bases de tout organisme vivant, est la capacité à se reproduire. La reproduction permet à un organisme de former un nouvel organisme, tout en lui transmettant son information génétique (Mendel, 1865). Nous l'avons vu précédemment, l'ADN peut être la cible de modifications épigénétiques. Dans cette partie nous nous attacherons à comprendre comment la méthylation de l'ADN est transmise d'un organisme parent à sa descendance, que ce soit lors d'une reproduction sexuée ou asexuée.

Outre sa fonction dans la protection du génome, la méthylation de l'ADN a également d'autres fonctions, notamment dans l'hérédité.

Les premières études sur la transmission des marques épigénétiques à la descendance, réalisées sur des cellules de mammifères, ont conclu que les marques épigénétiques (méthylation de l'ADN) n'étaient pas transmises à la descendance. Progressivement des exceptions ont été découvertes, remettant en cause ce dogme initial (revu dans Rakyan & Beck, (2006)).

Chez les plantes, un mécanisme de reprogrammation existe, qui induirait une élimination partielle des marques épigénétiques d'une génération à l'autre (revue dans Kawashima & Berger, (2014)).

5.1. Multiplication asexuée

La reproduction asexuée ou clonale repose sur la mitose. Lors de la mitose, l'ADN est dupliqué. Puis intervient une division permettant à partir d'une cellule mère diploïde d'obtenir deux cellules filles diploïdes identiques, contenant le même matériel génétique (Criqui & Genschik, 2002). La transmission des marques épigénétiques peut être assurée d'une part par la maintenance de la méthylation en contexte CG et CHG et par la méthylation *de novo* en contexte CHH, comme décrit précédemment.

Pendant la reproduction asexuée, l'information épigénétique est donc transmise de façon stable. Tout comme des mutations génétiques peuvent survenir, du fait d'erreurs liées aux polymérases (Garcia-Diaz & Bebenek, 2007), des épimutations peuvent survenir, via des erreurs dans la restauration de la méthylation (van der Graaf *et al.*, 2015).

5.1.1. Cas du greffage

Les connaissances actuelles sur la transmission des marques épigénétiques via le greffage tendent à supposer que l'ADN des greffons n'est pas méthylé au même degré que celui des parents (Fraga *et al.*, 2002). Chez *Pinus radiata* par exemple, on sait que la succession de greffages abouti à une diminution de la méthylation de l'ADN (Fraga *et al.*, 2002).

5.1.2. Cas de la multiplication *in vitro*

La multiplication *in vitro* permet de régénérer des plantes entières à partir de fragment de tissus végétaux. Ce type de multiplication provoque un grand nombre de changements pour les vitro-plants au niveau morphologique, physiologique, de l'expression de leurs gènes mais également de leur épigénome (Joyce *et al.*, 2003).

Il a été montré chez *Sequoiadendron giganteum*, qu'un plant juvénile maintenu *in vitro* présente un taux de méthylation de son ADN supérieur à un plant greffé cultivé en extérieur (Monteuuis *et al.*,

2008). Deux études sur des micro-propagations *in vitro* d'*Acacia mangium* et *Castanea sativa*, tendent à montrer que le niveau de méthylation est supérieur chez les micro-propagules juvéniles, par rapport à ceux présentant un phénotype mature (Baurens *et al.*, 2004; Hasbún *et al.*, 2005).

5.2. Multiplication sexuée

La reproduction sexuée repose quant à elle sur le mécanisme de la méiose. La méiose consiste en une duplication de l'ADN, suivi de deux divisions cellulaires (Mercier *et al.*, 2015). Ces divisions successives permettent la formation des gamètes mâle et femelle. Les gamètes formés sont haploïdes, c'est-à-dire que pour chaque gène il n'y a qu'un seul allèle.

5.2.1. Gamétogénèse mâle

Au sein des tissus reproducteurs mâles, va être formé le pollen mature (ou gamétophyte mâle, Kawashima & Berger, (2014)). Le pollen mature est composé d'une cellule végétative contenant deux gamètes (Figure 5.2-1). La cellule végétative permet la fécondation du gamétophyte femelle par la formation du tube pollinique.

Au cours de la formation du pollen mature, se produit un phénomène de reprogrammation épigénétique permettant notamment l'expression de gènes impliqués dans la méiose (Walker *et al.*, 2018). Ce phénomène a notamment été mis en évidence par une augmentation de la transcription des TEs (Gehring *et al.*, 2009; Slotkin *et al.*, 2009), du fait d'une diminution de la méthylation dans la cellule végétative (Gehring *et al.*, 2009; Hsieh *et al.*, 2009), mais pas dans les gamètes. Cette diminution de la méthylation est notamment possible grâce à l'action de DEMETER (DME) qui a une activité de déméthylation de l'ADN (Ibarra *et al.*, 2012). L'augmentation de la transcription des TEs au sein de la cellule végétative permettrait une production et un transfert de siRNA de la cellule végétative vers les gamètes (Slotkin *et al.*, 2009), afin d'y renforcer la répression de la transcription des TEs. Cette régulation durant la formation du pollen est nommée l'empreinte paternelle (« paternal imprinting », Rodrigues & Zilberman, (2015)) et permet de conserver une répression forte des TEs (Ibarra *et al.*, 2012).

5.2.1. Gamétogénèse femelle

Au sein des tissus reproducteurs femelle vont se former les gamétophytes femelles (Kawashima & Berger, 2014). Un gamétophyte femelle mature est constitué de 8 noyaux formant 7 cellules (Figure 5.2-1). Au sein de ce gamétophyte femelle, nous allons nous intéresser plus particulièrement à la cellule centrale (comportant deux noyaux) et à la cellule œuf. De façon analogue au processus existant lors de la formation des gamètes mâles, il y a au sein de la cellule centrale un processus de déméthylation du génome provoquant une augmentation de l'activité de la voie de la RdDM. Il y a ainsi une accumulation de siRNA dans la cellule centrale (Mosher *et al.*, 2009), permettant de renforcer la répression de la transcription des TEs dans la cellule œuf est ainsi y créer une empreinte maternelle (Hsieh *et al.*, 2009).

5.2.2. Fécondation

Lors de la fécondation il y a en fait une double fécondation. L'un des gamètes mâle va féconder la cellule centrale, où il y aura alors deux exemplaires du génome maternel et un exemplaire du génome paternel. La cellule centrale fécondée forme ainsi l'albumen qui va ensuite accumuler des réserves nutritives lors de la maturation de la graine, afin de nourrir le futur embryon. Le second gamète mâle va, quant à lui, féconder la cellule œuf qui devient dès lors le zygote.

Concernant la méthylation, il se déroule ainsi un double phénomène pendant la reproduction sexuée ; d'une part une déméthylation partielle dans les gamétophyte mâle et femelle résultant en l'activation de la RdDM ; d'autre part, outre le transfert de siRNA issus de l'activation de la RdMD, une partie des marques de méthylation non supprimées est conservée, permettant le transfert de l'empreinte paternelle et maternelle à l'embryon (Law & Jacobsen, 2010). Par la suite, lors de la fécondation, se produit une forte activation des gènes impliqués dans la méthylation de l'ADN (voies de la maintenance de la méthylation et de la RdDM). Ce niveau d'expression est maintenu pendant la première étape de l'embryogénèse (Jullien *et al.*, 2012). Ceci permettant une répression des éléments transposables afin de maintenir l'intégrité du génome (Kawashima & Berger, 2014).

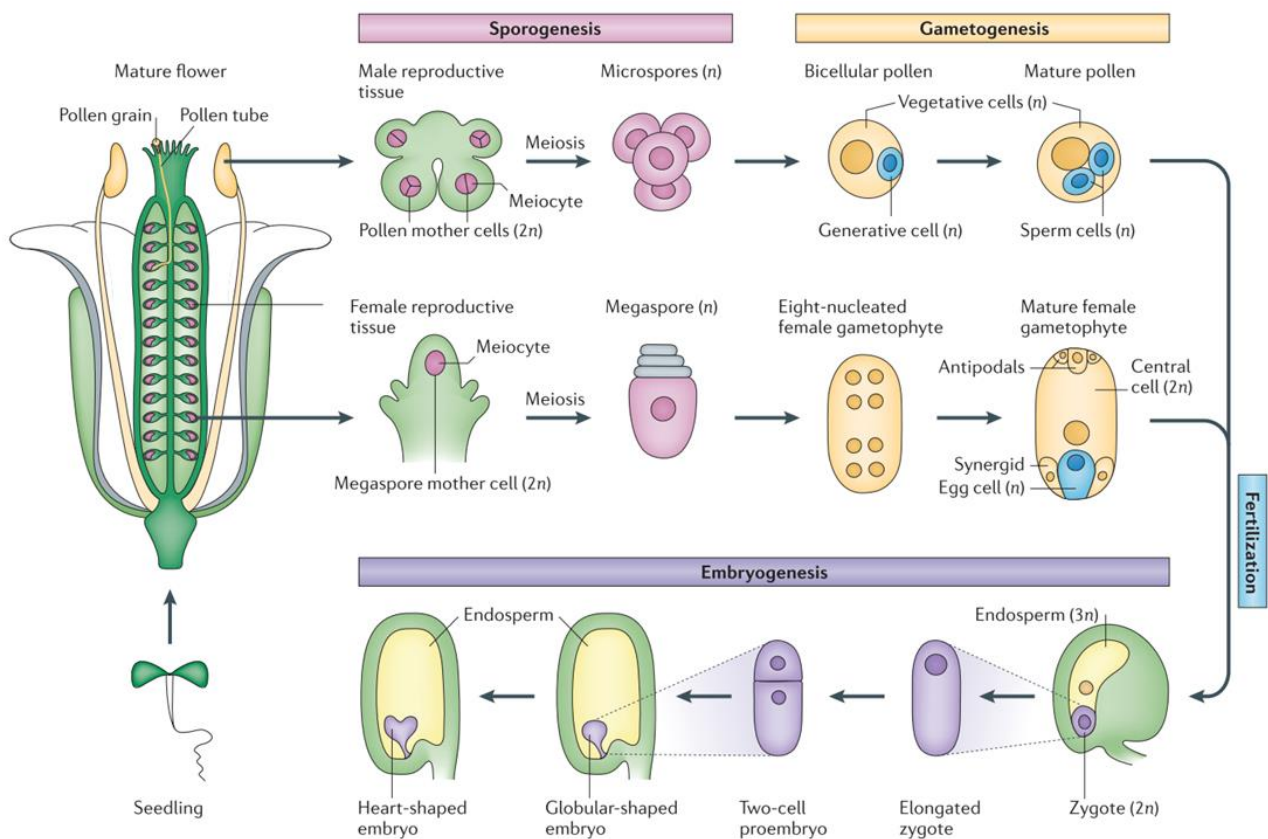


Figure 5.2-1: Schéma de la reproduction sexuée chez *A. thaliana* issue de la revue (Kawashima & Berger, 2014)

6. Objet de la thèse

6.1. Contexte

La transmission des marques épigénétiques par voie sexuée semble assurer la stabilité de l'expression de l'information génétique chez les espèces étudiées jusqu'à présent. Elle contribue sans doute également à une adaptation de ces espèces à un environnement donné (Richards *et al.*, 2010; Mirouze & Paszkowski, 2011; Casadesús & Low, 2013). Le pommier, comme d'autres espèces d'arbres, a un temps de génération long puisque sa période juvénile peut se prolonger au-delà de 8 ans (Visser, 1964). L'adaptation génétique à un stress engendré suite aux contraintes environnementales, via la multiplication sexuée, est donc plus lente pour ces arbres que pour des espèces annuelles, du fait de ce temps minimum entre deux générations. Nous pouvons donc supposer que les plantes pérennes vont privilégier la voie de l'acclimatation à leur écosystème, plutôt que la voie de l'adaptation génétique, pour se développer dans un environnement donné. Cette acclimatation, aussi nommé plasticité, est liée à la modification de l'expression des gènes et peut être induite notamment par des modifications épigénétiques (Herman & Sultan, 2011; Kinoshita & Seki, 2014; Xu *et al.*, 2018b).

Les pommiers sont des plantes pérennes principalement propagés par greffage (Fallahi *et al.*, 2002), afin de conserver une homogénéité variétale. Chez *Pinus radiata*, une autre espèce pérenne, il a été montré que des propagations par greffages successifs entraînent une diminution du niveau de méthylation (Fraga *et al.*, 2002). Il est donc possible que chez le pommier, du fait de son mode de propagation privilégié en verger, l'on puisse s'attendre à une diminution de la méthylation de l'ADN entre un arbre adultes mature et un jeune greffon qui en est issu.

6.2. Objectif

La question principale traitée dans le cadre de cette thèse est la suivante : Quelles sont les conséquences phénotypiques, transcriptomiques et épigénétiques engendrées par la multiplication sexuée et la propagation asexuée du pommier ?

De manière à répondre à cette question, nous allons étudier les différences entre un mode de multiplication sexué et un mode de reproduction asexué chez le pommier. Nous allons pour cela réaliser un phénotypage, un séquençage du méthylome et une analyse microarray du transcriptome d'un pommier adulte du génotype 'GDDH13', ainsi que de jeunes greffons et de semis obtenus par autofécondation de ce même génotype.

Une seconde question abordée dans le cadre de cette thèse est de savoir si certains régulateurs de la méthylation connus chez *A. thaliana* fonctionnent de manière similaire chez le pommier.

Les travaux menés pour répondre à cette seconde question s'intègrent au projet Genius (Genome ENgineering Improvement for Useful plants of a Sustainable agriculture). Nous allons tout d'abord chercher à comprendre comment la méthylation des cytosines est modifiée par la mutation du gène *METHYLTRANSFERASE1* du pommier, connu pour être impliqué dans la maintenance de la méthylation en contexte CG chez *A. thaliana*. Nous allons également tester si le gène *AtDRM2* peut permettre de modifier l'épigénome du pommier. Nous utiliserons pour cela la technique CRISPR-Cas9.

Chapitre I : Différences épigénétiques entre une multiplication sexuée et une propagation asexuée chez le pommier

1. Introduction

Durant ces dernières décennies, l'épigénétique est devenue un sujet d'étude majeur chez les plantes. De nombreuses études épigénétiques réalisées sur les plantes s'intéressent à l'effet d'un stress biotique (Downen *et al.*, 2012b; Yu *et al.*, 2013; Le *et al.*, 2014) ou abiotique (Chen *et al.*, 2010; Luo *et al.*, 2012; González *et al.*, 2013; Bocchini *et al.*, 2015) sur l'épigénome des plantes. Il est également régulièrement question d'une mémoire épigénétique d'un stress dans la descendance d'une plante (Boyko *et al.*, 2010; Slaughter *et al.*, 2012; van der Graaf *et al.*, 2015).

Il existe également des études analysant la diversité épigénétique des populations de plantes annuelles (Hagmann *et al.*, 2015; Kawakatsu *et al.*, 2016; Zhang *et al.*, 2018) ou pérennes (Salmon *et al.*, 2005; Keyte *et al.*, 2006; Avramidou *et al.*, 2015; Herrera & Bazaga, 2016) ayant évolué dans des environnements différents. Une étude réalisée sur 34 espèces d'angiospermes annuelles et pérennes différentes présentait une analyse de l'évolution de la méthylation chez les angiospermes (Niederhuth *et al.*, 2016). De cette étude, Niederhuth *et al.*, (2016) montrent une variabilité importante de la structure des méthylomes chez les angiosperme. Chaque méthylome étant spécifique d'une espèce et reflétant l'histoire de l'espèce. D'autres études se sont également intéressées à la structure du méthylome entre différentes espèces (e.g. Seymour *et al.*, (2014)). Seymour *et al.*, (2014) rapportent que certaines régions du génome sont hautement variables au niveau de leur méthylome, sans pour autant qu'il y ait un lien clair avec une variabilité phénotypique. En 2019, Albaladejo *et al.*, ont rapporté une étude sur le méthylome d'un pistachier mère par rapport à sa descendance issue d'une multiplication sexuée par pollinisation croisée. Ils ont ainsi montré une diminution du taux de méthylation dans la descendance mais également un lien au niveau de méthylation et stade de développement de la plante. Ils ont également montré qu'une forte méthylation du génome semble corrélérer avec une diminution du rendement.

Cependant sur l'ensemble de ces études, je n'ai pas identifié d'études s'intéressant à la comparaison d'une multiplication sexuée par rapport à une propagation asexuée de type greffage au niveau du méthylome d'une espèce.

Dans ce premier chapitre il sera donc question de comprendre comment se transmettent les marques épigénétiques chez une espèce pérenne appartenant à la famille des Rosaceae. Nous avons choisi de travailler sur le modèle pommier (*Malus domestica*) en utilisant un matériel unique : un pommier homozygote (GDDH13, Lespinasse *et al.*, (1999)) qui peut s'autoféconder. Nous avons ainsi pu profiter du fait que les plants issus de semis ou de greffage, obtenus à partir de cet arbre, sont génétiquement identiques. Un séquençage bisulfite de ces différents échantillons a été réalisé avant le début de ma thèse. Les séquences obtenues ont été alignées sur le génome du pommier 'GDDH13' récemment séquencé à l'IRHS d'Angers (Daccord *et al.*, 2017). Nous avons ensuite calculé le taux de méthylation de chaque cytosine afin d'évaluer ensuite les régions différentiellement méthylées (DMRs). Ce qui nous a permis par la suite d'étudier les portions du génome où la méthylation est conservée au cours des générations et quelles portions subissent des modifications en fonction du type de la multiplication utilisée pour produire l'échantillon (multiplication sexuée ou asexuée). Nous avons également étudié le phénotype de nos échantillons et réalisé une étude du transcriptome. Enfin nous avons cherché une corrélation entre les données de séquençage bisulfite avec les données transcriptomiques. Ces analyses ont fait l'objet d'une publication présentée dans la première partie de ce chapitre.

Dans une seconde partie nous avons cherché à analyser différentes méthodes de calcul de DMRs afin de définir quelle méthode serait la plus à même de décrire le profil de méthylation de notre modèle biologique

Pour l'article ci-dessous et la suite du manuscrit j'ai utilisé les termes **Old grafted tree (OG)**, **Young Graft (YG)** et **Seedling (SD)** pour nommer les échantillons. Les abréviations suivantes seront utilisées pour parler de la comparaison entre les différents individus. Ainsi la comparaison entre les **arbres adultes (OG)** et les **plants greffés (YG)** sera identifiée par **OGvsYG**. La comparaison entre les **arbres adultes (OG)** et les **plants issus de semis (SD)** sera identifiée par **SDvsOG**. Et enfin la comparaison entre les **plants greffés (YG)** et les plants issus de **semis (SD)** sera identifiée par **SDvsYG**.

2. Matériels et méthodes spécifiques

2.1. Matériel végétal

Nous avons utilisé un pommier homozygote, le génotype ‘GDDH13’, pour l’obtention de plants greffés et de plants issus de semis tous génétiquement identiques (Perrin *et al.*, 2020).

Le pommier présente une auto-incompatibilité pollinique (Zhang *et al.*, 1988). Cette caractéristique à plusieurs conséquences. Tout d’abord d’un point de vue génétique, le pommier présente une hétérozygotie importante (Li *et al.*, 2016). Ceci rend les analyses génétiques et épigénétiques plus complexes car pour chaque gène on trouve deux allèles différents pouvant également présenter des niveaux de méthylation différents. Ainsi il n’est pas aisé de différencier un effet génétique d’un effet épigénétique. De plus il n’est pas possible de séparer les épi-allèles lors du séquençage du méthylome. Enfin pour multiplier une variété de pommier inscrite au catalogue officiel on ne peut pas passer par la reproduction sexuée car il n’y aurait pas conservation des caractéristiques de la variété. Les pépiniéristes greffent donc les pommiers sur des portes greffes, permettant de conserver l’homogénéité des variétés. Cette étape de greffe est comparable à de la multiplication asexuée, il n’y a donc pas d’étape de méiose.

En serre nous avons à notre disposition un arbre adulte du génotype GDDH13 qui a été utilisé comme arbre parent de référence (OG). Des baguettes du génotype GDDH13 ont été prélevées dans les vergers sur 2 arbres clones de l’arbre de référence. Les greffes (YG) une fois réalisées ont été placées dans la même serre que l’arbre de référence. Les semis (SD) ont été réalisés à partir de pépins récoltés sur 5 clones, de l’arbre de référence du génotype GDDH13, placés sous cages avec une ruche de bourdons pour obtenir une autofécondation. Une fois germés et suffisamment développés les semis ont été transférés dans la même serre que l’arbre de référence.

Tous nos échantillons ont donc ainsi été placés dans les mêmes conditions de température, d’ensoleillement et d’humidité.

2.2. Séquençage bisulfite

L’analyse du méthylome nécessite de pouvoir le séquencer. Il existe pour cela une technique : le séquençage bisulfite. Cette technique permet de convertir toutes les cytosines de l’ADN qui ne sont pas méthylées en uracile (Figure 2.2-1). Les cytosines méthylées ne sont pas modifiées durant cette conversion. Avant d’être séquencé, l’ADN est amplifié pour augmenter sa quantité. Lors de l’amplification par PCR les uraciles présents sur un brin d’ADN sont remplacés par une thymine. Ainsi lors du séquençage les cytosines qui étaient méthylées seront détectées comme des cytosines, alors que les cytosines non méthylées auront été remplacées par des thymines. Les séquences ainsi obtenues sont alignées sur le génome.

La conversion bisulfite n’est pas parfaite malgré une efficacité supérieure à 99%, ainsi certaines cytosines non méthylées peuvent rester des cytosines au lieu d’être converties en thymines. De plus d’une cellule à l’autre l’ADN n’est pas toujours méthylé exactement de la même façon malgré les mécanismes de maintenance. C’est pourquoi à la suite du séquençage bisulfite et de l’alignement des séquences sur le génome de référence on définit un taux de méthylation pour chaque cytosine. Cependant les séquences présentant des taux de méthylation différents de 0% ou 100% sont très minoritaires dans le génome comme a pu le montrer Nicolas Daccord pendant sa thèse (Daccord, 2018).

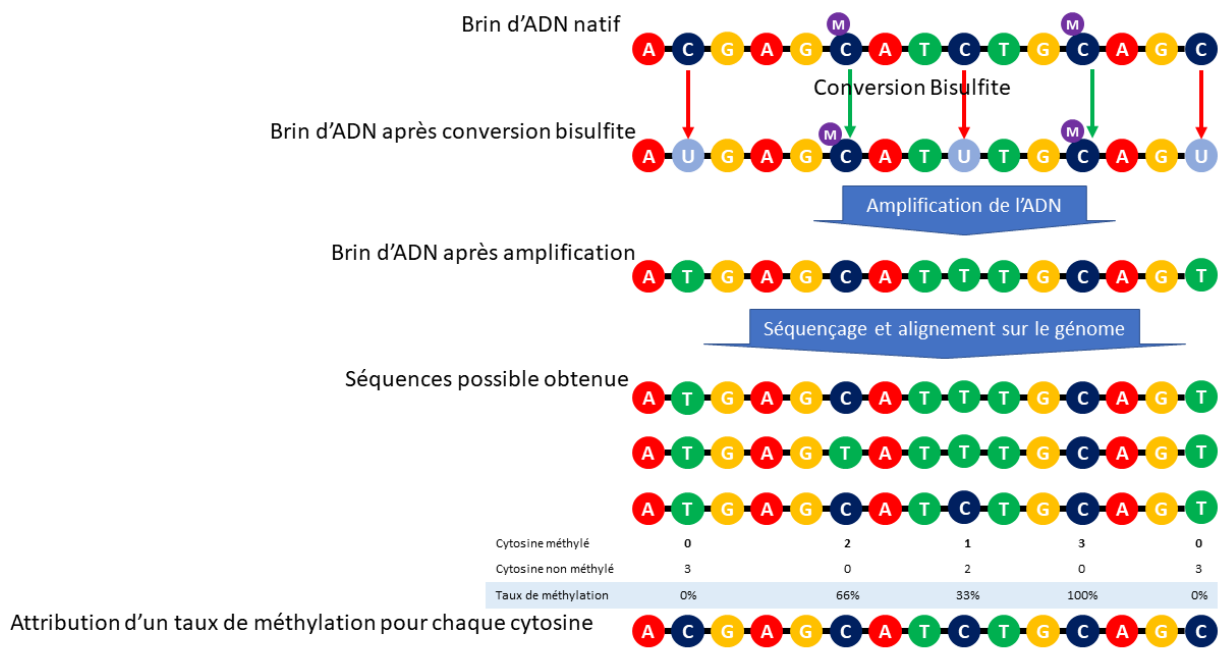


Figure 2.2-1 : schéma des étapes d'un séquençage bisulfite

3. Partie 1 – “Divergent DNA methylation signatures of juvenile seedlings, grafts and adult apple trees” (Perrin et al. 2020)

3.1. Introduction à l'article

L'article suivant a été publié le 01.03.2020 dans la revue Epigenomes (MDPI).

Nous avons ici publié une partie des résultats obtenues lors de nos analyses de l'hérédité des marques épigénétiques.

Nous avons notamment présenté une analyse phénotypique, transcriptomique et épigénétique réalisée sur des prélèvements issus de semis et de greffons par rapport à un arbre adulte.

Les figures supplémentaires à l'article sont disponibles en annexe, les données brutes sont disponibles sur la base de donnée GEO du NCBI, les calculs des régions différemment méthylé et des gènes différemment transcrits sont disponibles en suivant le lien indiqué dans l'article vers le site du journal.

3.2. Divergent DNA methylation signatures of juvenile seedlings, grafts and adult apple trees

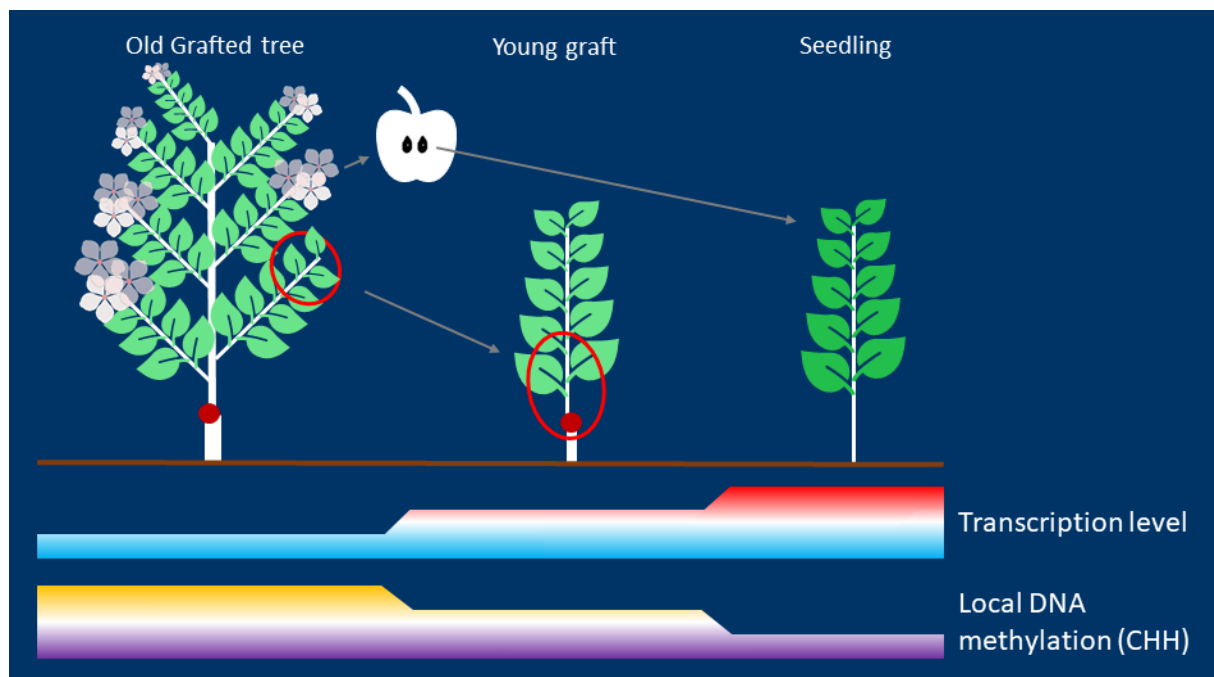




Figure 3.2-1: Graphical abstract of (Perrin et al., 2020) publication

Article

Divergent DNA Methylation Signatures of Juvenile Seedlings, Grafts and Adult Apple Trees

Adrien Perrin ^{1,*}, Nicolas Daccord ¹, David Roquis ^{1,2} , Jean-Marc Celton ^{1,*}, Emilie Vergne ^{1,*} and Etienne Bucher ^{1,2,*} 

¹ IRHS (Institut de Recherche en Horticulture et Semences), UMR 1345, INRA, Agrocampus-Ouest, Université d'Angers, SFR 4207 QuaSaV, F-49071 Beaucouzé, France; nicolas.daccord@gmail.com (N.D.); david.roquis@agroscope.admin.ch (D.R.)

² Plant Breeding and Genetic Resources, Agroscope, 1260 Nyon, Switzerland

* Correspondence: adrien.perrin@inra.fr (A.P.); jean-marc.celton@inra.fr (J.-M.C.); emilie.vergne@inra.fr (E.V.); etienne.bucher@agroscope.admin.ch (E.B.)

Received: 9 January 2020; Accepted: 27 February 2020; Published: 1 March 2020



Abstract: The vast majority of previous studies on epigenetics in plants have centered on the study of inheritance of DNA methylation patterns in annual plants. In contrast, perennial plants may have the ability to accumulate changes in DNA methylation patterns over numerous years. However, currently little is known about long-lived perennial and clonally reproducing plants that may have evolved different DNA methylation inheritance mechanisms as compared to annual plants. To study the transmission of DNA methylation patterns in a perennial plant, we used apple (*Malus domestica*) as a model plant. First, we investigated the inheritance of DNA methylation patterns during sexual reproduction in apple by comparing DNA methylation patterns of mature trees to juvenile seedlings resulting from selfing. While we did not observe a drastic genome-wide change in DNA methylation levels, we found clear variations in DNA methylation patterns localized in regions enriched for genes involved in photosynthesis. Using transcriptomics, we also observed that genes involved in this pathway were overexpressed in seedlings. To assess how DNA methylation patterns are transmitted during clonal propagation we then compared global DNA methylation of a newly grafted tree to its mature donor tree. We identified significant, albeit weak DNA methylation changes resulting from grafting. Overall, we found that a majority of DNA methylation patterns from the mature donor tree are transmitted to newly grafted plants, however with detectable specific local differences. Both the epigenomic and transcriptomic data indicate that grafted plants are at an intermediate phase between an adult tree and seedling and inherit part of the epigenomic history of their donor tree.

Keywords: epigenetics; perennial plant; transmission of methylation signatures; *Malus domestica*; sexual and asexual propagation

1. Introduction

Epigenetic regulation of gene transcription is implemented by several covalent modifications occurring at the histone or DNA level without affecting the DNA sequence itself [1]. These modifications are termed epigenetic marks and can change throughout plant development. Some newly acquired epigenetic changes can also be inherited across generations [2–5]. During their lifetime organisms may develop alternative phenotypes in response to biotic and abiotic stresses [6–9]. These stimuli result in modifications in gene transcription which can be altered by epigenetic modifications [10–12]. Besides gene transcription changes, certain epigenetic marks have been shown to play key roles in DNA conformation and genome stability [2,12,13]. Indeed, DNA methylation has been shown to have

a major role in transposable element (TE) silencing by reducing considerably the potential damage incurred by de novo TE insertions in the genome [14–16].

At the molecular level, DNA methylation consists in the covalent addition of a methyl group to cytosine nucleotide. In plants, DNA methylation occurs in three different cytosine contexts: CG, CHG and CHH (H= A, T or C) [17–20]. DNA methylation is established de novo or maintained by several DNA methyltransferase enzymes [21], each having a specific role depending on the sequence context. In order to maintain DNA methylation following DNA replication that results in hemi-methylated DNA, the methyltransferases MET1 and CMT3 can copy DNA methylation patterns from the “ancestral” strand to the newly synthesized strand. This mechanism is called DNA methylation maintenance [22,23] and occurs at symmetric CG and CHG sequence contexts. However, for the CHH sequence context no such template exists that may allow the DNA methylation maintenance mechanism. In this case, DNA methylation has to be restored by de novo methylation after each DNA replication cycle [24,25]. This pathway is called RNA-directed DNA methylation (RdDM) and requires small interfering RNAs (siRNA) [26,27] to guide the DNA methylation machinery to regions with sequence homology to the siRNAs.

From an epigenetic point of view, perennial plants are of particular interest as they have the potential to accumulate epigenetic modifications throughout their lifetime and may pass this information to the next generation. In addition, in the Rosacea family [28] numerous crops and ornamental plants are multiplied by asexual multiplication via grafting. This is interesting because in addition to the long lifetime of these plants, asexual multiplication involves only mitotic cell divisions [29] and thus presumably increases the chances of transmission of acquired epigenetic marks. If that was the case, epimutations could be quite common in grafted perennial plants. In contrast, during sexual reproduction meiosis can result in epigenetic reprogramming and, therefore, to loss of acquired epigenetic marks [30–32]. In *Arabidopsis*, this reprogramming is the result of active DNA demethylation driven by DEMETER (DME) [30]. Previous studies have suggested that this demethylation process could contribute to the generation of totipotent cells [3,4,33] by alleviating gene silencing via active removal of DNA methylation. These modifications at the DNA methylation level are necessary for normal meiosis [34]. While the RdDM pathway remains active in the egg cell [35], in the central cell of the mature female gametophyte and in mature pollen sperm cells, RdDM activity is reduced [4]. This decrease releases the transcription of TEs, thus resulting in the production of siRNAs derived from those. These siRNAs have been reported to be transported into the egg cell [36] to silence homologous loci in the maternal and paternal genomes [4,36–39]. Based on these findings, one may assume that during sexual multiplication, meiosis would allow restoration of a basal DNA methylation level in these species, while mitosis during asexual multiplication would maintain acquired epimutations.

In plants, inheritance of epigenetic marks has been primarily investigated notably in annual plants. Several studies point to the existence of broad epigenetic variations throughout wild populations of perennial and annual plants [40–42]. Other studies have demonstrated that epigenomic plasticity can allow environmental stress adaptation and improve response to future stresses [43–46]. Furthermore, it has been suggested that epigenetic modifications induced by stress in a mother plant may improve stress response in their offspring [47–49]. However, little is still known about heritable transmission of epigenetic marks in crops and more specifically in woody perennials like apple [29,50,51].

Apple (*Malus domestica*) is a major fruit crop in the world. In 2017, around 130 million tons of fruit were produced on approximately 12.3 million hectares [52]. Because apple is an obligate out crosser, tree propagation for commercial orchards and conservation has to be performed via asexual multiplication to maintain the traits of interest. This vegetative multiplication (or clonal multiplication) ensures that all grafted trees originating from a particular cultivar are largely genetically identical. Scions of fruiting cultivars are grafted on rootstock to combine valuable agricultural traits. For instance, in addition to reducing tree size and modifying its architecture, grafting onto particular rootstocks is known to shorten the juvenile phase of the scion by promoting flower differentiation [53]. Scions can thus recover their ability to bloom within 3 to 5 years after grafting [53] while seedlings on their

own roots may only start blooming after up to 8 years [54]. The length of the initial juvenile phase [55] can be highly variable among species, ranging from a few days, as in the *Rosa* genus [56] to more than 30 years in some woody plants [57–59]. Certain phenotypic characteristics have been associated with the juvenile phase such as fast vegetative growth [59], low lignification of young shoots, short internodes, specific leaf shape [55] and low trichome density. For instance, this phenotypic difference between juveniles and adults has previously been described in annual plants such as *Arabidopsis* [60] or *Zea mays* [61], and perennials including the *Acacia* genus, *Eucalyptus globulus*, *Hedera helix*, *Quercus acutissima* [62] or in *Populus trichocarpa* [63].

Here, to better understand how epigenetic marks at the DNA methylation level are transmitted in annual plants, we compared sexual (seed) and asexual (grafting) propagation in apple. Taking advantage of unique homozygous and self-compatible genetic material, we studied how these modes of propagation affect the phenotype, gene transcription and DNA methylation patterns. We present evidence that globally, genome-wide DNA methylation levels are stable in apple independent of the mode of multiplication. However, specific local variations in DNA methylation patterns were detected. These patterns were associated with the regulation of key plant-specific gene regulatory networks such as photosynthesis. This work provides a basis for future studies on the role of epigenetics in tree aging.

2. Results

2.1. Phenotypic Comparison of Seedlings, Young Grafts and Adult Trees

Unlike most heterozygous apple varieties, the Golden Delicious-derived GDDH13 doubled haploid apple shows a high self-compatibility level. To prevent outcrossing and to produce self-fertilized GDDH13 seeds, we covered trees with insect- and wind-proof cages during blooming time. Then we deployed bumblebees into the cages resulting in the production of hundreds of self-fertilized seeds. This unique material allowed us to study genetically identical seedlings and grafted plantlets derived from the very same parental tree. For that purpose, we simultaneously planted seedlings and grafted budwood from GDDH13 to ensure that the growing plants were of comparable size.

First, we studied the phenotypic differences between parental tree, grafts and seedlings on leaf samples in order to assess whether the plants were in a juvenile or adult phase. Trichome density was the most noticeable phenotypic difference (Figure 1). Leaves sampled from seedlings (SD) displayed a notably lower trichome density on their abaxial face (Figure 1A) compared to the original parental tree (old graft, OG). Leaves sampled from young grafted plants (YG) displayed a similar trichome density as OG (Figure 1B–D).

2.2. Transcriptional Profiles of Seedlings, Young Grafts and Adult Trees

In order to identify genes related to the juvenile phenotype or genes displaying differential transcription levels in response to grafting, we performed a set of differential gene transcription analyses. We assessed steady state RNA levels by performing the following two comparisons: OG versus SD (OGvsSD) and OG versus YG (OGvsYG). Transcriptomes were obtained using a custom-designed microarray that includes probes from all annotated GDDH13 genes and a fraction of transposable elements (TEs). We identified 6943 and 7353 differentially expressed transcripts (DETs) for OGvsSD and OGvsYG, respectively. Of these DETs, 5695 were annotated as genes (differentially transcribed gene, DTGs) in OGvsSD and 4996 in OGvsYG (Figure 2A). In total these DTGs include 13.5% of all annotated genes on the microarray for OGvsSD and 11.8% for OGvsYG (Figure 2A). For transcripts annotated as TEs, we identified 1248 and 2357 differentially transcribed transposable elements (DTTEs) in the OGvsSD and OGvsYG comparisons, respectively (Figure 2B). These represent 5% of all annotated TEs on the microarray for OGvsSD and 6.6% for OGvsYG (Figure 2B).

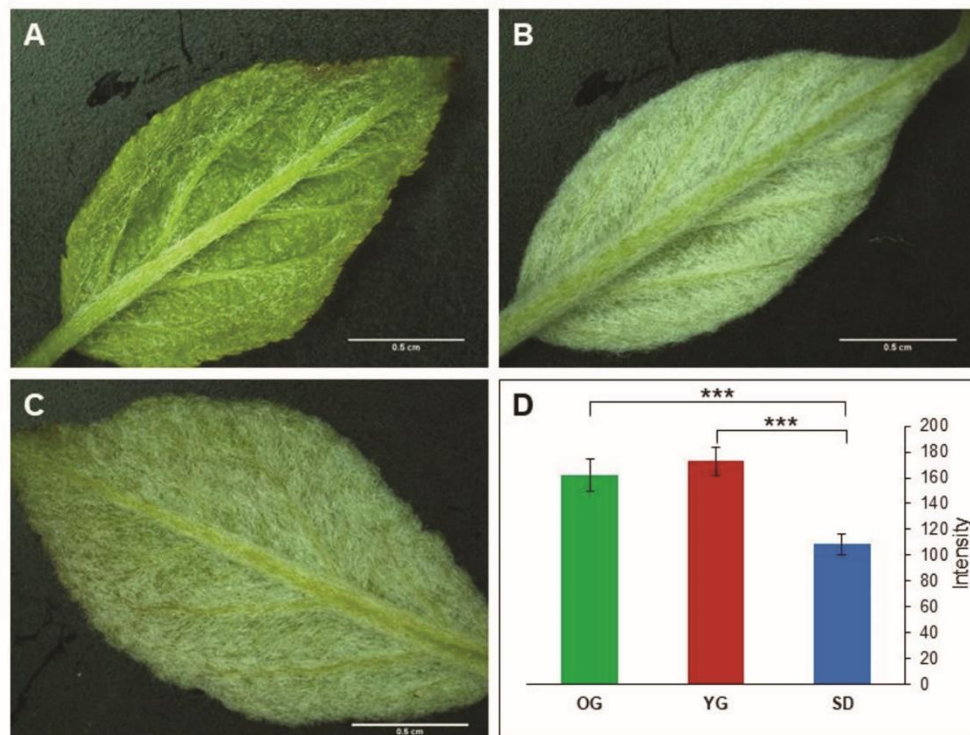


Figure 1. Leaf trichome density comparisons between seedlings (SD), young grafted plants (YG) and old graft (OG). Leaf pictures indicate visual differences in trichome density for SD (A), YG (B) and donor OG (C). The graph in (D), represents results from light intensity measurements carried out on the abaxial face of leaves. High light intensity correlates with high trichome density. $N = 60$ (5 measurements on 12 leaves) per sample. Statistical differences were evaluated by a two-by-two Kruskal–Wallis test. Asterix p -value: *** 1%.

Overall, DTGs displayed a tendency to be up-regulated in SD and YG compared to OG (Figure 2A). However, for TEs only the OGvsYG comparison followed the same pattern, since up- and down-regulated TEs were more equally distributed in the common DTTEs group. DTTEs specific to OGvsSD displayed a tendency to be down-regulated in SD.

Focusing on the common DTGs between OGvsSD and OGvsYG, we observed two groups (Figure 2A,C). The first group is composed of the 2.085 DTGs displaying a similar regulation pattern: 1365 and 720 DTGs were down- and up-regulated in OGvsSD and OGvsYG, respectively. In the second smaller group, only 85 DTGs displayed an opposite trend: these transcripts were up-regulated in SD in OGvsSD, but down-regulated in YG in OGvsYG. Similarly, we observed two groups for DTTEs (Figure 2B,D). 277 DTTEs were down-regulated in SD and YG in both OGvsSD and OGvsYG respectively, and 226 DTTEs were up-regulated in SD and YG in both comparisons compared to OG. Only 17 DTTEs displayed opposite transcript accumulation patterns compared to the general trend.

To study the main gene regulatory pathways represented in the differential transcription data we used the GDDH13 gene annotation of *Malus domestica* (v1.1) combined with the MapMan software ([64], Figure 3A). We also considered the TE class repartition as previously described in Daccord et al., [65] (Figure 3B). We observed variations in class sizes between OGvsSD and OGvsYG. The most notable variations in size were observed for: photosynthesis (9% of variation in total DTGs in OGvsSD and only 1% in OGvsYG), cell cycle (2% in OGvsSD and 9% in OGvsYG), solute transport (9% in OGvsSD and 4% in OGvsYG), cytoskeleton (1% in OGvsSD and 6% in OGvsYG), RNA biosynthesis (13% in OGvsSD and 18% in OGvsYG), RNA processing (2% in OGvsSD and 6% in OGvsYG) and chromatin organization (2% in OGvsSD and 5% in OGvsYG).

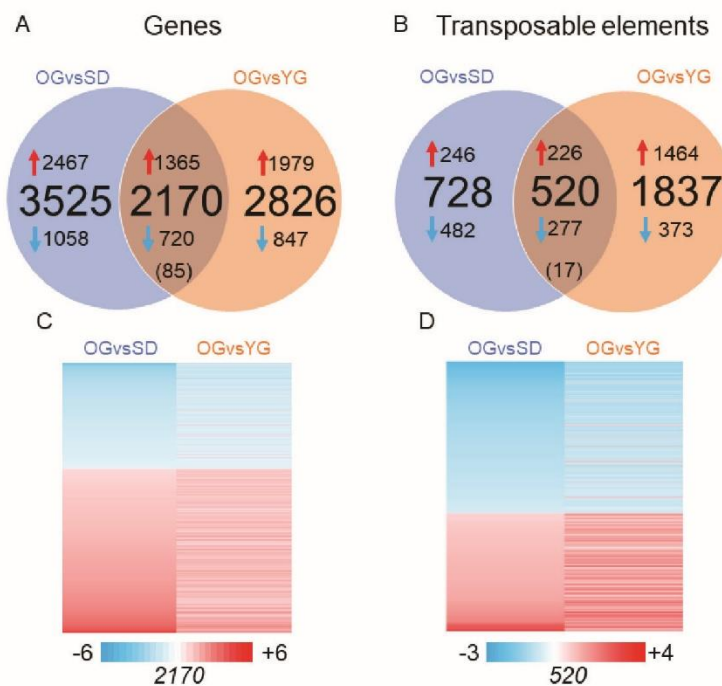


Figure 2. Transcriptome comparisons between SD, YG and OG. Graphical representation of the number of differentially expressed transcripts in the different comparisons. **(A)** Venn diagram showing differentially transcribed genes (DTGs) in the comparisons OGvsSD and OGvsYG. **(B)** Venn diagram depicting differentially transcribed transposable elements (DTTEs) in the comparisons OGvsSD and OGvsYG. The central number in brackets represent common transcripts displaying opposed patterns of transcriptional regulation. In **(C)** and **(D)** the heat maps depict transcription ratios of common DTGs **(C)** and DTTEs **(D)**. Blue arrow depicting transcripts down regulated in SD and YG compared to OG and red arrow depicting transcript up regulated in SD and YG compared to OG. Numbers of DETs in each heat map are indicated below it. Fold change ratios are shown in the color scale bar.

In order to identify overrepresented classes of genes that could be linked to either the adult or the juvenile phase, we performed an enrichment analysis with the MapMan software using our DTGs as input data (Table S6). In the OGvsSD comparison, seven functional categories were overrepresented including coenzyme metabolism, terpenoids metabolism, chromatin organization, squamosa binding protein (SBP) family transcription factor, protein biosynthesis, peptide tagging in protein degradation and enzyme classification. Eleven classes are overrepresented in the OGvsYG comparison (Table S6), including secondary metabolism, chromatin organization, cell cycle, RNA processing, protein biosynthesis, peptide tagging, cytoskeleton, cell wall, solute transport, and enzyme classification.

Next, we considered the TE class repartition in our DTTE list (Figure 3B). We did not find large variations in the class repartition among comparisons. Class I TE represented 53% of DTTEs on the microarray in OGvsSD and 46% in OGvsYG. Concerning class II TEs we found 31% and 43% of DTTEs in OGvsSD and in OGvsYG, respectively.

Altogether, our analyses show that the sexual and asexual tree propagation methods investigated here had a significant effect on gene and TE transcription in GDDH13.

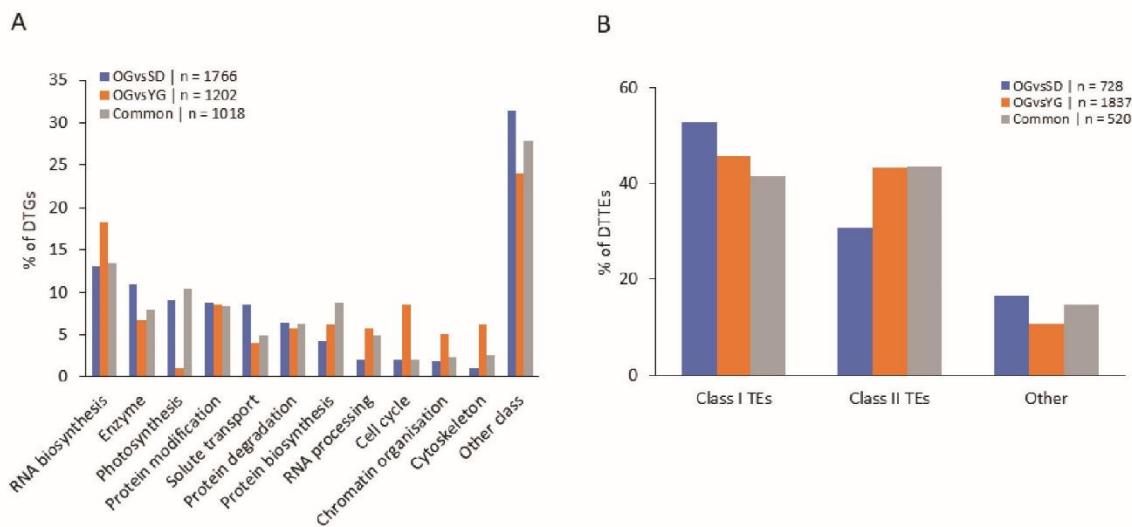


Figure 3. Classification of differentially expressed transcripts. **(A)** Percentage of DTGs in each comparison in function of the gene classification according to Lohse et al. [64]. **(B)** Percentage of DTTEs in each comparison in function of TEs classification according to Daccord et al. [65]. Classes represented by less than 5% in the three conditions were summed up in “Other class”. Transposon in “other” class correspond to potential host gene and unclassified TEs according to Daccord et al. [65].

2.3. Global DNA Methylation Analysis of Seedlings, Young Grafts and Adult Trees

To investigate how DNA methylation marks are transmitted through mitosis as compared to meiosis, we assessed DNA methylation levels in SD, YG and OG samples at the genome-wide level by using whole genome bisulfite sequencing (WGBS). First, we compared genome-wide DNA methylation levels at cytosines in the three sequence contexts (CG, CHG, CHH). Our primary investigation indicated that there was no significant difference in cytosine methylation averages, in any of the contexts among the tested samples (Figure 4A).

Next, we computed and identified differentially methylated regions (DMR) between SD, YG and OG. Overall, we identified 99,437 DMRs in OGvsSD and 44,001 in OGvsYG (Figure 4B). We also investigated DMRs close to genes (Gene-DMRs) or TEs (TE-DMRs). These DMRs are defined by their relative proximity to genes or TEs. For this purpose, we selected DMRs located within 2000 bp 5′ or 3′ of annotated genes or TEs as was previously done in Daccord et al. [65]. We identified 30,902 and 12,150 Gene-DMRs in OGvsSD and OGvsYG, respectively. For TEs, we identified 42,143 and 21,909 TE-DMRs in OGvsSD and OGvsYG, respectively (Figure 4B).

We found that in each comparison, in genes, TEs or other genomic loci, DMRs were largely hypomethylated in SD and YG compared to OG (Figure 4B). Indeed 94% and 63% of DMRs in the three contexts were hypomethylated in SD and in YG respectively compared to OG. Moreover, a vast majority of DMRs were identified in the CHH context (99% and 97% in OGvsSD and OGvsYG, respectively; Table 1). Overall, DMRs in the CHH context tended to be hypomethylated in SD and YG compared to OG (95% in OGvsSD and 66% in OGvsYG) and hypermethylated in SD and YG compared to OG in the CG and CHG contexts (85% in OGvsSD and 93% in OGvsYG) (Table 1). To identify whether DMRs were equally distributed along the genome, or were regrouped within hot spots, we computed the DMR density for the individual contexts as shown in Figure 4C. Overall, we found DMRs to be equally distributed all along the apple chromosomes, with some regions displaying a higher enrichment (Figure 4E, red boxes).

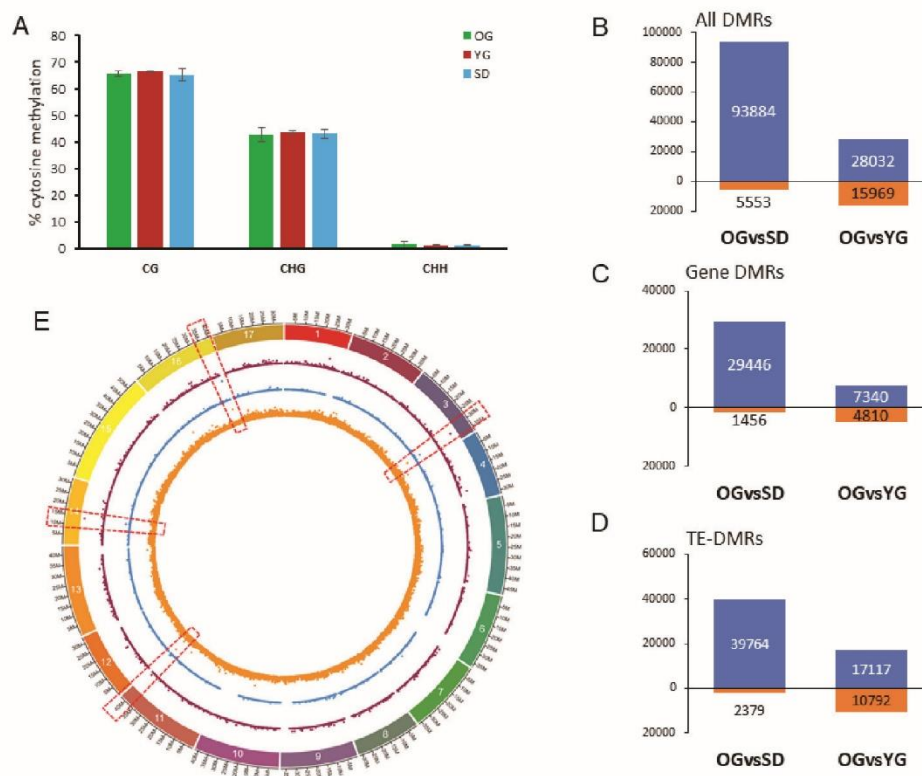


Figure 4. Global overview of DNA methylation differences between SD, YG and OG. (A) Clustered column chart presenting the genome wide cytosine methylation level (in percentage) of the three methylation contexts (CG, CHG and CHH). Student test was performed to evaluate differences but none was found to be significant. (B–D) stacked graph representing the number of differentially methylated regions (DMRs) for each comparison: hypomethylated (above 0, in blue) or hypermethylated (below 0, in orange) in the SD and YG samples compared to OG for all DMRs (B), Gene-DMRs (C) and TE-DMRs (D). DMRs in all sequence contexts were counted and values are indicated in graph. (E) Density plot of number of DMRs in 50 kb windows on the GDDH13 genome for OGvsSD (see Figure S1 for OGvsYG). In violet, DMRs in the CG context, in blue for the CHG context and in orange the CHH context. Each point represents the number of DMRs in a 50kb window of the genome. Red dashed boxes indicate the presence of DMR hot spots.

Table 1. DMR distributions according to context and methylation changes. Number and percentage of hypo- and hypermethylated DMRs in SD and YG sample compared to OG in each comparison (OGvsSD and OGvsYG). Numbers indicated in grey correspond to the percentage of DMRs in the specific cytosine context in the comparisons. Numbers given in blue indicate the percentage of all DMRs in the corresponding comparison.

Context	Old Graft vs. Seedling			Old Graft vs. Young Graft		
	Hypomethylated	Hypermethylated	Σ	Hypomethylated	Hypermethylated	Σ
	Number (%)	Number (%)	Number (%)	Number (%)	Number (%)	Number (%)
CHH	93.784 (94,9)	4998 (5,1)	98.782 (99,3)	27.945 (65,5)	14.749 (34,5)	42.694 (97,0)
CHG	22 (12,3)	157 (87,7)	179 (0,2)	19 (4,9)	370 (95,1)	389 (0,9)
CG	78 (16,4)	398 (83,6)	476 (0,5)	68 (7,4)	850 (92,6)	918 (2,1)
Σ	93.884 (94,4)	5553 (5,6)	99.437	28.032 (63,7)	15.969 (36,3)	44.001

In order to quantify and compare DNA methylation levels we quantified DNA methylation changes (δmC) within DMRs in each sequence context (Figure 5). Overall, we identified significant differences in δmC for the CHG and CHH and not for the CG sequence contexts. Interestingly, in the CHG context, the δmC value was higher in OGvsYG (12.1%) than in OGvsSD (3.6%) for hypomethylated DMRs in YG and SD compared to OG. In the CHH context, we observed that the δmC value was higher in OGvsSD (7.0%) than in OGvsYG (6.1%) for hypomethylated DMRs in YG and SD compared to OG, and lower in OGvsSD (3.7%) than in OGvsYG (7.5%) for hypermethylated DMRs in YG and SD compared to OG. From these results, we conclude that the transmission of cytosine methylation from OG to SD is different to that from OG to YG depending on the cytosine sequence context.

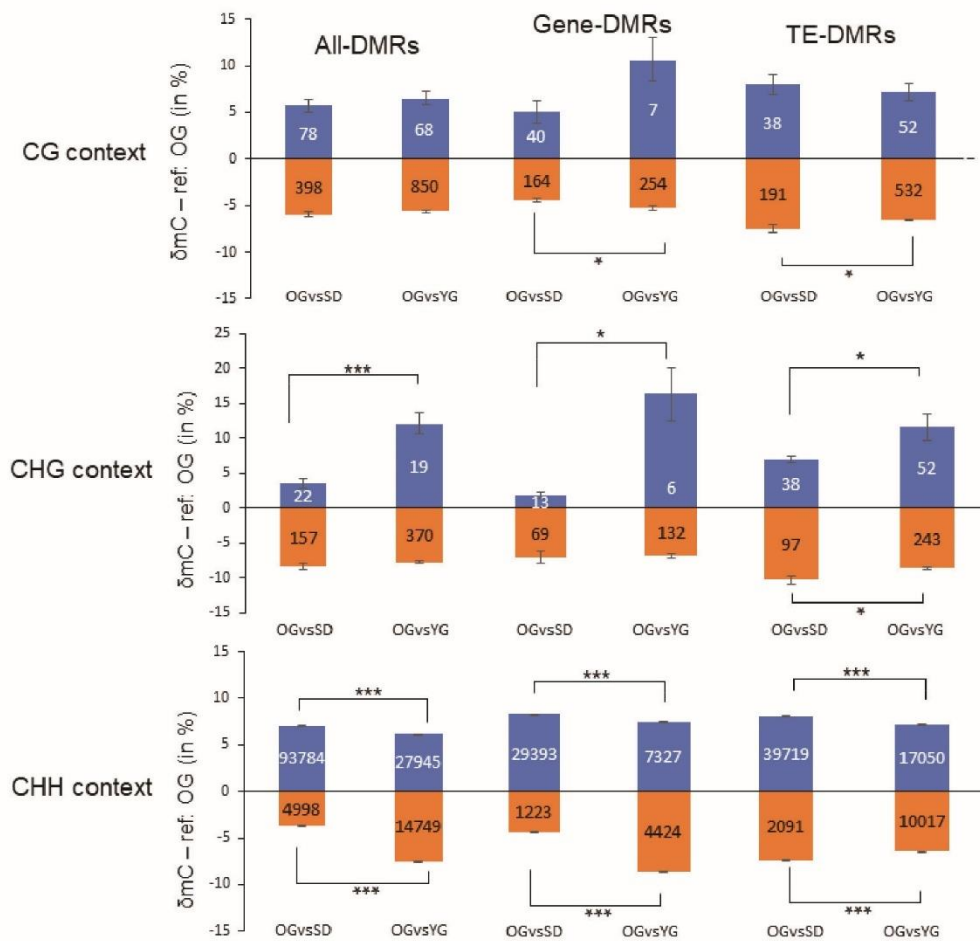


Figure 5. Levels of DNA methylation changes in gene and TE annotations. Stacked graph depicting DMR methylation variations (δmC) between samples separated by sequence context and functional annotation. All DMRs taken together are presented in the All-DMRs column, DMRs at genes and TEs in the Gene-DMRs and TE-DMRs columns, respectively. DMRs were filtered by p -value and SDA (standard deviation average) in accordance to a fixed threshold (Table S2). Student test was performed to evaluate differences in δmC , results are represented by an asterisk depending on the p -value threshold: * 5%; ** 1%; *** 1%. δmC : delta of methylation. In blue and orange, the hypomethylated and hypermethylated DMRs in SD and YG compared to OG, respectively.

For DMRs located in genic regions (Gene-DMRs, Figure 5) we observed that there were less DMRs in the CG-CHG (286 for OGvsSD and 393 for OGvsYG) contexts than in the CHH context (30,616 for OGvsSD and 11,751 for OGvsYG). Gene-DMRs in the CG and CHG contexts were almost all hypermethylated in YG and SD compared to OG in both comparisons. Indeed, 80% and 99% of Gene-DMRs in the CG context were hypermethylated in OGvsSD and OGvsYG, respectively. 84% and

96% of Gene-DMRs were hypermethylated in CHG in OGvsSD and OGvsYG, respectively. This is consistent with the observations we made for the All-DMRs group (Table 1). Conversely, 96% and 62% of Gene-DMRs in the CHH context were hypomethylated in SD and YG compared to OG for OGvsSD and OGvsYG, respectively.

While studying the DNA methylation changes, we found that in the CG and CHH contexts, the δmC values of hypomethylated Gene-DMRs were smaller in OGvsSD (4.2% for both comparisons) than in OGvsYG (4.9% and 8.2%, respectively). However, for hypomethylated Gene-DMRs in the CHG context in YG or SD compared to OG, the δmC value was higher in OGvsYG (16.3%) than in OGvsSD (1.8%) following the overall trend observed for All-DMRs. Then, for hypomethylated Gene-DMRs in the CHH context in YG or SD compared to OG, the δmC value was higher in OGvsSD (7.7%) than in OGvsYG (6.9%) following the trend observed for All-DMRs. These observations indicate towards a contrasted sequence context-specific pattern of DNA methylation differences.

For DMRs located in TE annotations (TE-DMRs, Figure 5), our observations were similar to the results for Gene-DMRs. Overall, most TE-DMRs were hypermethylated in YG or SD compared to OG in the CG (83% for OGvsSD and 91% for OGvsYG) and CHG (72% for OGvsSD and 82.4% for OGvsYG) contexts, and hypomethylated in the CHH context (95% for OGvsSD and 63% for OGvsYG). As for Gene-DMRs we found only differences to hypermethylated TE-DMRs in CG context, but here, δmC value is higher in OGvsSD (7%) than in OGvsYG (6.1%). For CHG, the δmC value of hypomethylated TE-DMRs was smaller in OGvsSD (6.9%) than in OGvsYG (11.6%) and higher for hypermethylated TE-DMRs in OGvsSD (10.3%) as opposed to OGvsYG (8.6%). Finally, in CHH context, we observed that δmC value was higher in hypomethylated and hypermethylated TE-DMRs in OGvsSD (7.5% and 7%, respectively) than in OGvsYG (6.7% and 6.1%, respectively).

Overall, even though there were no large global differences in DNA methylation level between the samples analyzed here, we found significant local differences. The majority of DMRs were in the CHH context with a tendency to be hypomethylated in YG and SD compared to OG.

2.4. Classes of Genes Enriched with Differentially Methylated Regions (DMRs)

To identify genes belonging to particular functional categories and presenting DMRs in their proximity, we used the aforementioned GDDH13 annotation in MapMan and the TE annotation as previously used in our transcriptomic analysis. Here, we only considered Gene-DMRs and TE-DMRs in the CHH context. We excluded DMRs associated with the CG and CHG context from further analyses due of their very limited number (Tables S2 and S3). For the following, we termed as DTG-DMRs genes that we found to be differentially transcribed and containing or being close to DMRs. Similarly, TEs identified as DTTEs and being associated with TE-DMRs were termed DTTE-DMRs.

As expected, we found the seven classes that we previously identified in DTGs analysis: RNA biosynthesis, protein modification, enzyme family, protein degradation, solute transport, photosynthesis and protein biosynthesis (Figure 6A). We did not find differences in the proportion of gene classes between OGvsSD and OGvsYG.

For DTTE-DMRs (Figure 6C) we observed a smaller proportion of Class I TEs in OGvsYG (57.4%) compared to OGvsSD (77.7%), while for class II TEs we found 35.4% for OGvsYG and 15.5% for OGvsSD.

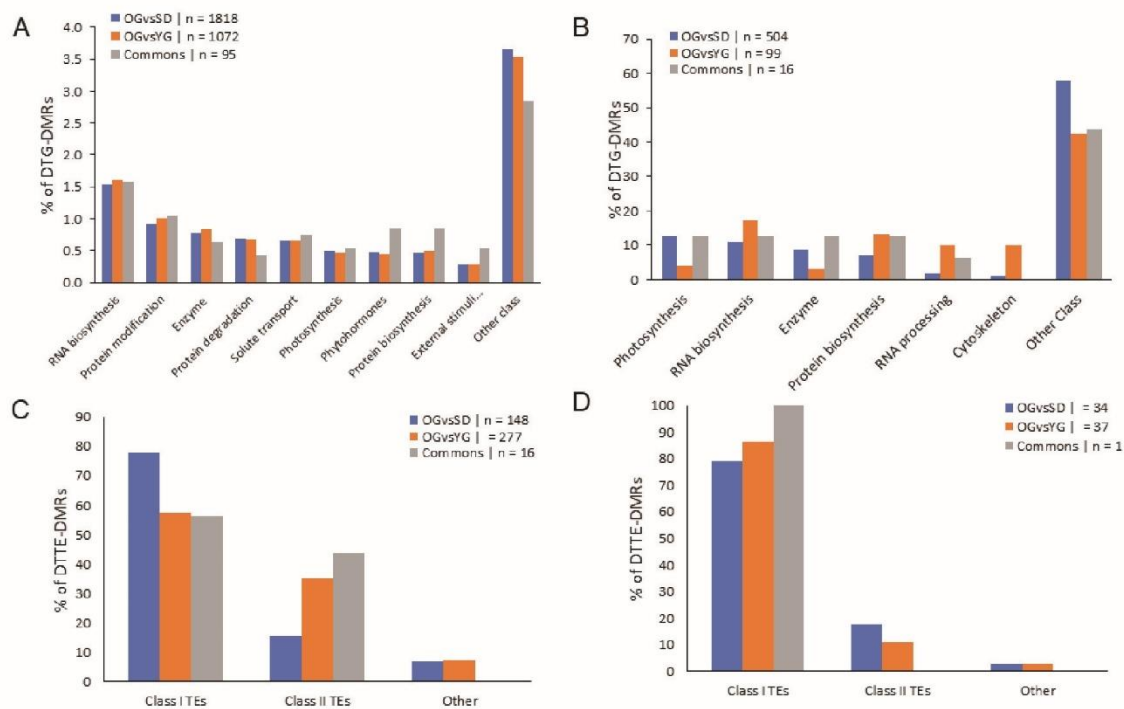


Figure 6. Classification of differentially transcribed genes that are associated to DMRs. Clustered column charts depicting the percentage of DTG-DMRs (A and B) and DTTE-DMRs (C and D) in the respective comparisons in function of gene or TEs classification. Only DMRs in the CHH context are presented here. In (A) and (C) all DTGs- and DTTEs-DMRs were used while in (B) and (D) we only considered DTGs and DTTEs with differential transcription ratio greater than 1.5 in absolute value. Gene classes representing less than 5% (A) or 10% (B) of the total in the three conditions were summed up in “other class”.

2.5. Relationship between DNA Methylation and Transcription

In order to identify correlations between DNA methylation and gene expression we associated Gene- and TE-DMRs to our microarray transcriptome data. The gene classes were defined according to the Mapman annotation of genes and to the TE annotation previously used to analyze DTGs and DTTEs.

We found 520 DTG-DMRs in OGvsSD and 115 DTG-DMRs in OGvsYG (Figure 6C), 35 DTTE-DMRs in OGvsSD and 38 DTTE-DMRs in OGvsYG (Figure 6D). We investigated genes and TEs classes' repartition for DTG- and DTTE-DMRs. Of the 11 classes found in DTGs analysis (Figure 3A), we found six gene classes representing only slightly more than 5% of all DTG-DMRs. These included the classes photosynthesis, RNA biosynthesis, enzyme family, protein biosynthesis, RNA processing and cytoskeleton (Figure 6B).

We did not observe notable shifts within the classes' repartition between OGvsSD and OGvsYG for DTTE-DMRs (Figure 6D).

Finally, we investigated the link between DMRs and DTGs. Again, we only considered DTG-DMR in the CHH context due to the low number of DTG-DMR we found in the CG and CHG contexts. We noticed that in both comparisons the majority of DMRs in DTG-DMRs were located in gene promoters (553 for OGvsSD and 126 for OGvsYG), followed by the terminator region (284 for OGvsSD and 71 for OGvsYG) finally followed by those present in gene bodies (111 for OGvsSD and 26 for OGvsYG) (Figure 7A,B. See examples of these DTG-DMRs in Figure S2).

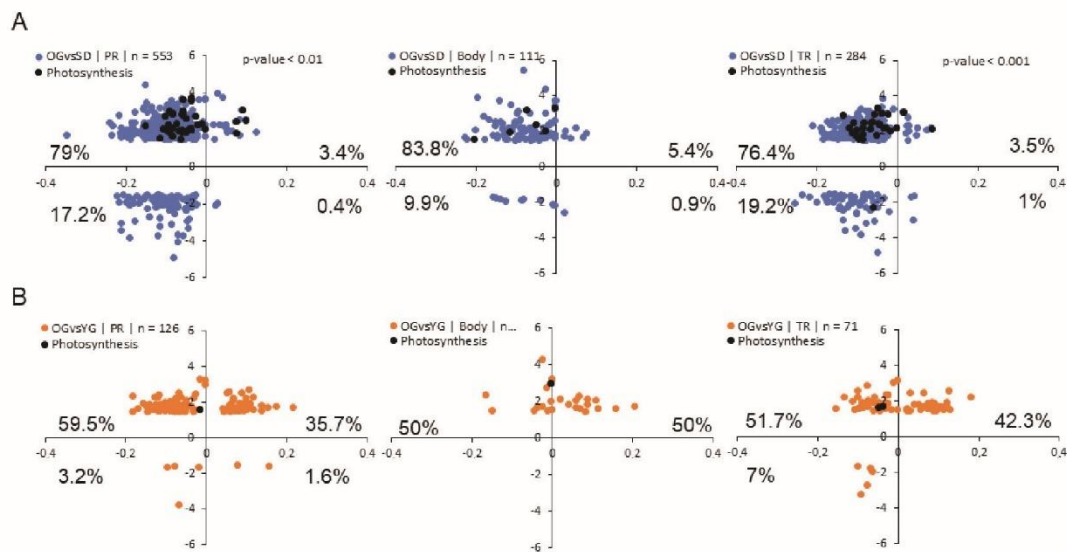


Figure 7. Relationship between transcription ratio and DNA methylation variation Scatterplot representing DTG-DMRs in OGvsSD (A) and OGvsYG (B) in the CHH context, X axis represents δmC and the Y axis represents gene expression ratios of the corresponding comparisons. In blue/orange are shown all DTG-DMRs and in black the ones specifically associated to photosynthesis. Numbers of DTG-DMRs used in each graph are indicated in the legend and corresponding percentages in each quadrant at the edges of the graph. We separated DTG-DMRs in function of the position of the DMRs related to the corresponding gene (PR = promoter, body, TR = terminator). Student test was performed to evaluate the correlation between expression and DMR state, significant correlations are indicated with p -values on corresponding scatterplot.

Our data also indicates that, hypomethylated DMRs located in promoter and terminator regions of genes, were associated with transcriptional up-regulation in SD compared to OG (Figure 7A) but we did not find that correlation in the OGvsYG comparison.

Among the classes of DTGs associated with DMRs, we observed that genes associated with photosynthesis tended to be both hypomethylated and up-regulated in SD or YG compared to OG. Indeed 92% of Gene-DMRs associated with the photosynthesis pathway presented this pattern in OGvsSD and 100% of them in OGvsYG.

Our results indicate that in the CHH context, hypermethylation of a DNA sequence in the proximity of a gene correlates with a reduced level of transcription of that particular gene following sexual propagation.

3. Discussion

3.1. Newly Grafted Plants Are at an Intermediate State between Adult Tree and Juvenile Seedling

Phenotypic differences between juvenile and adult plants are commonly observed at the leaf level [55]. In our study we observed that leaves of apple seedlings displayed a low trichome density compared to the donor tree (Figure 1). In contrast, young grafted plants did not display a difference in trichome density compared to the donor tree. As previously reported by others, this phenotype can be associated to the juvenile phase [66] and a young grafted plant is thus developmentally closer to the adult tree than to the juvenile seedling. Nonetheless, newly grafted plants show a contrasted ability to flower. In the grafting process, a mature bud (able to flower or quiescent) is placed on a short-rooted stem (rootstock). The number of nodes between the apical bud and the rootstock is drastically reduced to 1 or 2 nodes. After their first year of growth, buds are in a mature adult state but are unable to initiate flowers and to bear fruits because of an insufficient number of nodes (less than 77) in the stem.

This limit was previously described as a transition phase between juvenile and adult apple tree [67–69]. Thus, young grafted plants are not adult plants from a physiological point of view.

Here, we wanted to study the molecular changes that occur during propagation via grafting and by seed formation. First, we compared the transcription profiles in three different stages: seedlings (SD), young grafted plants (YG) and an adult tree (old graft, OG), taking advantage of our genetically identical material growing under controlled conditions. Globally we observed a higher transcription level for the majority of the DTGs and DTTEs in the SD or YG compared to OG. This correlates well with the previously reported decrease in gene transcription in mature compared to juvenile plants [70–72]. Furthermore, the common DTGs identified as activated in the OGvsSD and OGvsYG comparisons could be correlated to high vegetative growth in younger stages. Such rapid growth can be observed in seedlings and in newly grafted plants, and in mature apple trees gene expression is generally reduced, as observed in Day et al. [73]. Thus, the transcriptome of newly grafted plants showed similarities with the one obtained from seedlings but also with that of adult trees. These observations are in line with previous studies on other woody plants [70–73].

Overall, our findings indicate that young grafted plants are at the interface between a juvenile seedling and an adult mature tree [67,68] from a morphological and transcriptomic perspective.

This intermediate condition of newly grafted plants is also confirmed from a molecular point of view. Indeed, we identified differences in gene class repartition for DTGs between OGvsSD and OGvsYG (Figure 3A) which included photosynthesis, RNA processing, chromatin organization and cell cycle classes. As regards genes related to photosynthesis, we found that they represented 9% of the DTGs in OGvsSD but only 1% in OGvsYG. This is consistent with previous observations showing that the photosynthetic pathway is differentially regulated between juvenile and mature reproductive plant, especially in woody plants (reviewed in [74]). Indeed, photosynthesis is known as a physiological process subjected to many modifications during the transition from the juvenile to the mature phase [75]. As juveniles, SD undergo broader transcriptomic changes compared to YG and OG. This can be associated to an age-related gene transcription pattern previously described for photosynthesis-related genes in other woody plant such as *Pinus taeda* [76], *Larix laricina* [77], *Picea rubens* [78] and in the *Quercus* genus [79].

This intermediate condition of newly grafted plants between juvenile seedlings and adult tree was also observed at specific loci at the DNA methylation level in the CHH context. Indeed, overall a hypomethylation of the CHH-DMRs was observed in SD and YG compared to OG. We also found a less extensive hypomethylation in OGvsYG (65.5% of all DMRs were hypomethylated in YG compared to OG) than in OGvsSD (95% of all DMRs were hypomethylated in SD compared to OG).

3.2. DMRs Correlate with Neighboring Gene Transcription

Previous reports established a correlation between DNA methylation and the repression of gene transcription, particularly in *Arabidopsis* [80,81]. In this study, we investigated a possible link between DNA methylation and gene transcription changes in *M. domestica*. For that purpose, we associated DMRs with their neighboring DTG in order to investigate the effect of methylation on gene transcription. We found that in the CHH context, genes with closely located hypomethylated DMRs in promoter or terminator regions in SDs compared to OG often also displayed a higher transcription level (Figure 7). This was particularly the case for photosynthesis-related genes (Figure 7). Our correlative observations thus indicate that cytosine methylation in the CHH context may to be involved in regulating the transcription of these genes.

We did not only observe local changes in DNA methylation, but also contrasted levels of DNA methylation changes (δmC). Indeed, we found significant differences at the δmC level between both comparisons, particularly in the CHG and CHH contexts (Figure 5). For methylation in the CHH context, we observed that, even if the difference in δmC was significant between OGvsSD and OGvsYG, it was not very high between the comparisons but also within comparisons. Indeed, the highest δmC was on average above 8% for hypermethylated Gene-DMRs in OGvsYG. But we also found that

this relatively small methylation variation was sufficient to relate it with gene transcription changes (Figure 7).

4. Materials and Methods

4.1. Plant Material

Malus domestica plants were obtained from GDDH13 [82] line (X9273). Young grafted plants (YG) were obtained by grafting budwood of GDDH13 orchard tree (planted in 2001) on the rootstock cultivar 'MM106'. Seedlings (SD) were obtained by self-fertilizing GDDH13 flowers on an isolated tree located in INRAE's orchards. Seed dormancy was removed after a 3 months period of cold stratification before sowing. The homozygous state of the seedlings used in this study was evaluated using publicly available SSR markers. A clone of the original GDDH13 tree, grafted onto an 'MM106' rootstock in 2007 and placed in the greenhouse in 2016 was used as reference mature adult tree (old graft, OG).

4.2. Phenotyping

Nine young leaves were collected for each sample (old graft, seedling and young graft) and time point. At each sampling time SD and YG plants were pruned to increase vigor. Three biological replicates were used at three weeks intervals for YG and SD materials in 2018, and one replicate was collected in 2019 including OG material (twelve leaves were sampled). Each leaf was then photographed using binocular magnifier (Olympus SZ61, Schott KL 1500 LED, Olympus DP20). Pictures were further analyzed with the ImageJ[®] software [83]. Pictures were transformed in 8-bit grayscale and light intensity was measured on 5 areas of 0.03 cm² on each leaf. Intensity differences among samples were evaluated using the R language by the Kruskal–Wallis test. We first analyzed biological replicates harvested in 2018 and 2019 (SD and YG). Because there were no differences among biological replicates of SD in 2018 and 2019, and similarly for YG sample in 2018 and 2019, we present only the results of 2019, which also include the OG sample.

4.3. DNA and RNA Extraction

The youngest and completely opened leaf below the apex was sampled for each replicate. Sampling was performed as described in Table S1. DNA was extracted using NucleoSpin Plant II kit (Macherey-Nagel, Hoerd, France). The manufacturer's recommendations were applied with the following modifications: at step 2a PL1 buffer volume was raised to 800 µL and PVP40 was added (3% of final volume), suspension was then incubated 30 min at 65 °C under agitation. The lysate solution was centrifuged 2 min at 11,000 g before transferring the supernatant in step 3. At step 4 PC buffer volume was raised to 900 µL. In step 6 the first wash volume was decreased to 600 µL and the third wash volume was raised to 300 µL. An extra-centrifuge step was added after washing to remove ethanol waste from the column. In step 7 DNA was eluted twice in 55 µL in total. The RNA was extracted using the NucleoSpin[®] RNA kit (Macherey-Nagel, Hoerd, France) according to the manufacturer's protocol.

4.4. Bisulfite Sequencing and DMRs Calling

Extracted DNA was precipitated using a mix of pure ethanol (70%), water (24%) and NaAc 3M (3%). After precipitation DNA was sent to Beijing Genomics Institute (Shenzhen, Guangdong 518083, China) in pure ethanol for whole genome bisulfite sequencing (WGBS). An average depth coverage of 14X per sample was obtained. DNA methylation data can be accessed on the Gene Expression Omnibus website under accession codes GSE138377. Bisulfite sequencing reads were mapped on GDDH13_V1.1 reference genome with Bsmapping tool [84] to obtain methylation calling file. Methylation averages between samples were compared by Student t-test using R [85].

We named differentially methylated regions (DMRs) using a pipeline developed in IRHS (Angers, France) published at GitHub platform (<https://github.com/EtienneBucher/DAMOCLES>). DMRs were

calculated between OG and SD samples and between OG and YG samples with the following parameters: minimum coverage of 3×, 200 bp sliding windows with 100 bp overlaps. DMRs files contain quality values such as *p*-value, standard deviation of averages (SDA) and methylation differences. SDA evaluates the natural variation between biological replicate, a high variation between replicates gives high SDA value. We empirically determined a threshold for each DNA methylation sequence context using visual inspections of DMR preview in JBrowse [86]. SDA thresholds were fixed as mentioned in Table S2. Thresholds were determined in order to select the most reproducible DMRs within biological replicates (Table S3; GSE138377).

In order to evaluate the variation of methylation levels within DMRs, we extracted DNA methylation level difference of YGvsOG and SDvsOG. Student test was performed to evaluate difference in methylation levels.

4.5. Microarray

The *Malus domestica* microarray (Agilent-085275_IRHS_Malus_domestica_v1; GPL25795; Agilent, Foster City, CA, USA) was used for transcriptome analysis. Details about this array are available on Gene Expression Omnibus at following link: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL25795>. Briefly, this array contains 89,218 coding DNA sequences (sense and antisense), 582 miRNAs (sense and antisense) and 13,834 ncRNAs (sense and antisense). This array also includes all annotated TEs families consisting in 69,966 transposable elements probes (TEs, sense and antisense). Complementary DNA (cDNA) were synthesized and hybridized with the Low Input Quick Amp Labeling Kit, two-color (Agilent, Foster City, CA, USA). Two biological replicates were used. Each biological replicate represents one sample for OG and YG materials, and a pool of two samples for SD material. Hybridizations were performed on a NimbleGen Hybridization System 4 (mix mode B) at 42 °C overnight. Slides were then washed, dried, and scanned at 2 μm resolution. NimbleGen MS 200 v1.2 software was used for microarray scans, and the Agilent Feature Extraction 11.5 software was used to extract pair-data files from the scanned images. We used the dye switch approach for statistical analysis as described in Depuydt et al. [87]. Analyses were performed using the R language (R Development Core Team, 2009); data were normalized with the lowess method, and differential transcription analyses were performed using the *lmFit* function and the Bayes moderated test using the package LIMMA [88]. Transcriptomic data are available in Gene Expression Omnibus website, with the accession GSE138491.

4.6. Reverse Transcriptase Quantitative Polymerase Chain Reaction (RT-qPCR) Microarray Validation

Extracted mRNA was treated by DNase with the RQ1 RNase-Free DNase (Promega, Madison, WI, US) following the manufacturer's protocol. The Moloney Murine Leukemia Virus Reverse Transcriptase was used to obtain cDNA from 1.2 μg of total RNA, with oligo(dT) primers following the manufacturer's protocol (Promega, Madison, WI, USA). For quantitative polymerase chain reaction (qPCR) measurements, 2.5 μL of cDNA at the appropriate dilution were mixed in a final volume of 10 μL with 5 μL of quantitative PCR mastermix (MasterMix Plus for SYBR Green I with fluorescein; Eurogentec EGT GROUP, Seraing, Belgium), with 0.2 μL of each primer (200 nM final) and with 4.1 μL of pure water. Primers were designed with Primer3Plus [89] and were used at their optimal concentration using the reaction efficiency calculation (near to 100%) according to Pfaffl recommendations [90]. Genes used to validate the microarray data were selected in differentially transcribed gene (DTG) lists in both comparisons (OGvSD and OGvYG) with 1) a high ratio value and 2) high intensities values. Accessions and primer sequences are indicated in Figure S3A. Reactions were performed with a CFX connect real-time system (Bio-Rad, Hercules, CA, USA) using the following program: 95 °C, 5 min; 35 cycles comprising 95 °C for 3 s, 60 °C for 45 s; 65 °C, 5 s and 90 °C for 1 min, with real-time fluorescence monitoring. Melting curves were acquired at end of each run. Data were acquired and analyzed with CFX Maestro V1.1 (Bio-Rad, Hercules, CA, USA). Gene transcription levels were calculated using the $2^{-\Delta\Delta C_t}$ method and were corrected as recommended by Vandesompele et al. [91] (Figure S3B), with three reference genes: *ACTIN* (accession CV151413, MD14G1142600), *GAPDH* (accession CN494000,

MD16G1111100), and *TUBULIN* (accession CO065788, MD03G1004400) used for the calculation of a normalization factor.

4.7. Differentially Expressed Transcript (DET) Analysis

Differentially expressed transcripts were selected based on their p -value $\leq 1\%$ (Table S4). For DET other than TEs and miRNA a MapMan annotation (<https://mapman.gabipd.org/home>; version 3.5.0BETA), was performed, using GDDH13_1-1_mercator4 map file, in order to assign each DET to a class of genes (BIN code). DET not assigned to a BIN class were excluded. A representativeness percentage of each BIN class was then calculated in the comparisons OGvSD, OGvYG and in the intersection between the both comparisons. A MapMan enrichment analysis using the BIN class representativeness was performed and a BH correction was applied [92] because of the high number of values. The DETs associated to a gene annotation are called DTGs in this manuscript. For differentially transcribed TEs (DTTEs), the TE classification [65] was used in order to assign each DTTE to a class.

4.8. Association between DMR and Differentially Transcribed Gene (DTG) or Differentially Transcribed Transposable Element (DTTE)

DMRs and transcription level data (DTG or DTTE) results were connected thanks to gene identifier. DMRs without associated DETs were removed. There is some redundancy in the gene or TE identification because several DMRs could be located to the same gene or TE. To avoid biases in our analysis we only kept DMRs with the highest methylation variation for each gene or TE (Table S5). For this analysis we applied a threshold and kept only transcripts with differential expression ratios above 1.5 and below -1.5 in order to better identified pathways or genes to work with. In order to identify correlation between DNA methylation and gene expression, we extracted the coefficient of determination of the linear regression of each scatterplot (Figure 7). We calculated the square root of the coefficient of determination and performed a Student test. Next, we compared this value to the theoretical value of the Student test. If a correlation was found, we indicated p -values on the scatterplot. Because of the low number of values ($n < 50$) in gene body localization in OGvsYG comparison, we first performed a Fisher test that validated the homogeneity of variance, thus allowing us to perform a Student test on these values.

5. Conclusions

In this study we compared the transmission of epigenetic marks and their potential effects on transcription during sexual and asexual propagation in apple.

First, we identified a phenotypic change (Figure 8) that was associated with adult plant phase and confirmed that grafting is not comparable to a complete rejuvenation process, as observed in seedlings. In our transcriptomic analysis we notably found that the transcription level of genes related to photosynthesis was elevated in seedlings compared to the tree, while newly grafted plants displayed an intermediate transcription level (Figure 8).

Analysis of the methylation data indicated that at the genome scale, the level of methylation in all three samples was similar. However, we were able to identify DMRs particularly in the CHH context. This result indicates that DNA methylation reprogramming during meiosis may not affect the global methylation level of the genome, but rather modifies particular regions of the genome. This observation was particularly striking regarding genes associated with photosynthesis. As found in the transcriptomic analysis, the DNA methylation data also indicated that grafted plants were at the interphase between the tree and the seedlings.

Globally, our results indicate that, from a physiological, transcriptomic and epigenomic standpoint, newly grafted plants are at the interphase between a tree and a seedling, displaying characteristics that are particular to both the mature and the young immature stages of the plant.

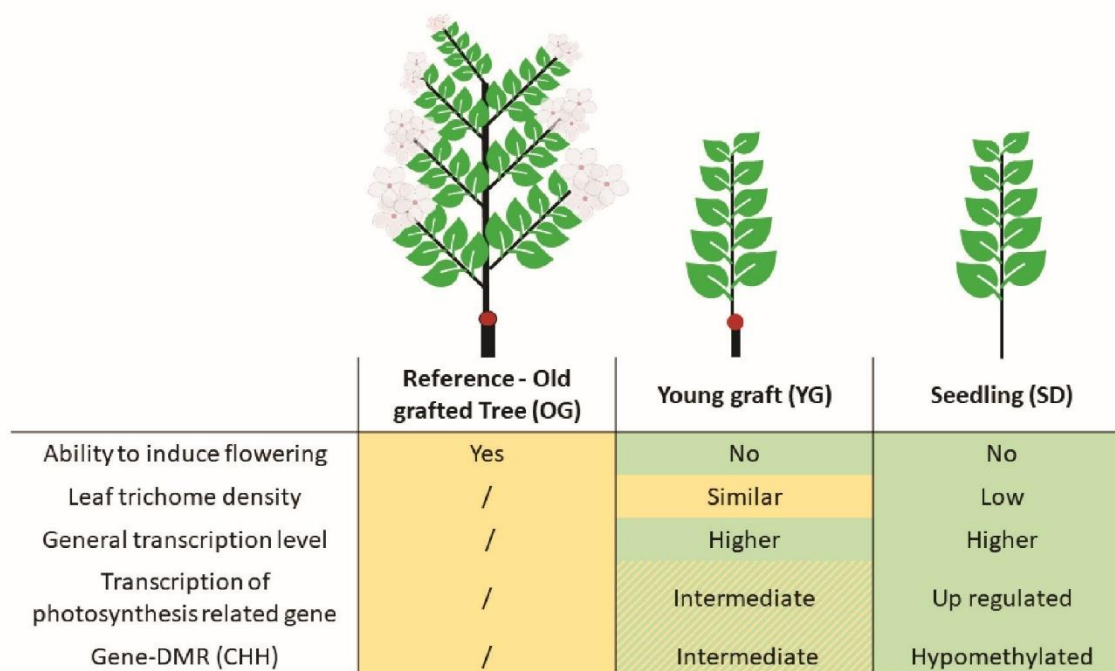


Figure 8. General overview of the main results of this study concerning physiological and molecular changes occurring during sexual and an asexual propagation. The red dot represents the grafting point between scion and rootstock (larger line weight). Shared aspects between plants are highlighted by the background colours.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2075-4655/4/1/4/s1>: Figure S1: Microarray validation, Figure S2: Methylation overview in GDDH13, Figure S3: Jbrowse screenshot, Table S1: Resume of defined samples and details of sampling, Table S2: Fixed threshold to filter DMRs calculated between each comparison, Table S3: DMRs list of OGvsSD and OGvsYG comparisons, including Gene- and TE-DMRs, Table S4: DETs list of OGvsSD and OGvsYG comparison, Table S5: DTG- and DTTE-DMRs list of OGvsSD and OGvsYG comparisons, Table S6: Functional classes found in enrichment analysis on Mapman software, Table S7: Count of DTG-DMRs in OGvsSD and OGvsYG comparisons, Table S8: Count of DTTE-DMRs in CHH context in OGvsSD and OGvsYG comparisons. Data Archiving Statement (secure tokens were provided to editor to allow reviewers' access). WGBS: methylation calling files, DMRs list and transcriptomic data were deposited on Gene Expression Omnibus online platform under the accession number GSE138492. We obtained two sub-accession numbers. The first, GSE138377, comprising all data related to methylation analysis as WGBS data, processed data and an excel spreadsheet of calculated and filtered DMRs to OGvsSD and OGvsYG comparison named Supp_Tab_S3.xlsx. The second, GSE138491, comprising the transcriptomic data obtained by microarray as describe in method.

Author Contributions: Conceptualization, J.-M.C. and E.B.; Formal analysis, A.P., N.D., D.R., J.-M.C., E.V. and E.B.; Investigation, A.P.; Methodology, E.V.; Software, N.D. and D.R.; Supervision, J.-M.C., E.V. and E.B.; Writing—original draft, A.P.; Writing—review and editing, A.P., J.-M.C., E.V. and E.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: M. Orsel-Baldwin for GDDH13_1-1_mercator4 files processing. Pays de la Loire to funding this work (project EPICENTER).

Conflicts of Interest: The authors declare no competing interest.

References

- Holliday, R.; Pugh, J.E. DNA Modification Mechanisms and Gene Activity during Development. *Source Sci. New Ser.* **1975**, *187*, 226–232. [[CrossRef](#)] [[PubMed](#)]
- Hauser, M.T.; Aufsatz, W.; Jonak, C.; Luschnig, C. Transgenerational epigenetic inheritance in plants. *Biochim. Biophys. Acta Gene Regul. Mech.* **2011**, *1809*, 459–468. [[CrossRef](#)]

3. Gutierrez-Marcos, J.F.; Dickinson, H.G. Epigenetic Reprogramming in Plant Reproductive Lineages. *Plant Cell Physiol.* **2012**, *53*, 817–823. [[CrossRef](#)] [[PubMed](#)]
4. Kawashima, T.; Berger, F. Epigenetic reprogramming in plant sexual reproduction. *Nat. Rev. Genet.* **2014**, *15*, 613–624. [[CrossRef](#)]
5. Quadrana, L.; Colot, V. Plant Transgenerational Epigenetics. *Annu. Rev. Genet.* **2016**, *50*, 467–491. [[CrossRef](#)]
6. Madlung, A.; Comai, L. The effect of stress on genome regulation and structure. *Ann. Bot.* **2004**, *94*, 481–495. [[CrossRef](#)]
7. Mirouze, M.; Paszkowski, J. Epigenetic contribution to stress adaptation in plants. *Curr. Opin. Plant Biol.* **2011**, *14*, 267–274. [[CrossRef](#)]
8. Köhler, C.; Wolff, P.; Spillane, C. Epigenetic Mechanisms Underlying Genomic Imprinting in Plants. *Annu. Rev. Plant Biol.* **2012**, *63*, 331–352. [[CrossRef](#)]
9. Song, J.; Irwin, J.; Dean, C. Minireview Remembering the Prolonged Cold of Winter. *Curr. Biol.* **2013**, *23*, 807–811. [[CrossRef](#)]
10. Manning, K.; Tör, M.; Poole, M.; Hong, Y.; Thompson, A.J.; King, G.J.; Giovannoni, J.J.; Seymour, G.B. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* **2006**, *38*, 948–952. [[CrossRef](#)]
11. 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. Patterns of population epigenomic diversity. *Nature* **2013**, *495*, 193–198. [[CrossRef](#)] [[PubMed](#)]
12. Kim, M.Y.; Zilberman, D. DNA methylation as a system of plant genomic immunity. *Trends Plant Sci.* **2014**, *19*, 320–326. [[CrossRef](#)] [[PubMed](#)]
13. Suzuki, M.M.; Bird, A. DNA methylation landscapes: Provocative insights from epigenomics. *Nat. Rev. Genet.* **2008**, *9*, 465–476. [[CrossRef](#)]
14. Miura, A.; Yonebayashi, S.; Watanabe, K.; Toyama, T.; Shimada, H.; Kakutani, T. Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. *Nature* **2001**, *411*, 212–214. [[CrossRef](#)]
15. 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. [[CrossRef](#)]
16. Ito, H.; Gaubert, H.; Bucher, E.; Mirouze, M.; Vaillant, I.; Paszkowski, J. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature* **2011**, *472*, 115–119. [[CrossRef](#)]
17. Gruenbaum, Y.; Naveh-Many, T.; Cedar, H.; Razin, A. Sequence specificity of methylation in higher plant DNA. *Nature* **1981**, *292*, 860–862. [[CrossRef](#)]
18. Meyer, P.; Niedenhof, I.; Ten Lohuis, M. Evidence for Cytosine Methylation of Non-Symmetrical Sequences in Transgenic Petunia Hybrida. *EMBO J.* **1994**, *13*, 2084–2088. [[CrossRef](#)]
19. Finnegan, E.J.; Genger, R.K.; Peacock, W.J.; Dennis, E.S. DNA Methylation in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1998**, *49*, 223–247. [[CrossRef](#)]
20. Chan, S.W.-L.; Henderson, I.R.; Jacobsen, S.E. Gardening The Genome: DNA Methylation in Arabidopsis thaliana. *Nat. Rev. Genet.* **2005**, *6*, 351. [[CrossRef](#)]
21. Law, J.A.; Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **2010**, *11*, 204–220. [[CrossRef](#)] [[PubMed](#)]
22. Lindroth, A.M. Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation. *Science (80-)* **2001**, *292*, 2077–2080. [[CrossRef](#)] [[PubMed](#)]
23. Schermelleh, L.; Haemmer, A.; Spada, F.; Rö Sing, N.; Meilinger, D.; Rothbauer, U.; Cardoso, M.C.; Leonhardt, H. Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.* **2007**, *35*, 4301–4312. [[CrossRef](#)] [[PubMed](#)]
24. Wassenegger, M.; Heimes, S.; Riedel, L.; Sängler, H.L. RNA-directed de novo methylation of genomic sequences in plants. *Cell* **1994**, *76*, 567–576. [[CrossRef](#)]
25. Chedin, F.; Lieber, M.R.; Hsieh, C.-L. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16916–16921. [[CrossRef](#)]
26. Herr, A.J.; Jensen, M.B.; Dalmay, T.; Baulcombe, D.C. RNA polymerase IV directs silencing of endogenous DNA. *Science (80)* **2005**, *308*, 118–120. [[CrossRef](#)]

27. Kanno, T.; Huettel, B.; Mette, F.; Aufsatz, W.; Jaligot, E.; Daxinger, L.; Kreil, D.P.; Matzke, M.; Matzke, A.J.M. Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat. Genet.* **2005**, *37*, 761–765. [[CrossRef](#)]
28. Hummer, K.E.; Janick, J. Rosaceae: Taxonomy, Economic Importance, Genomics. In *Genetics and Genomics of Rosaceae*; Springer: New York, NY, USA, 2009; pp. 1–17.
29. Verhoeven, K.J.F.; Preite, V. Epigenetic variation in asexually reproducing organisms. *Evolution (N. Y.)* **2014**, *68*, 644–655. [[CrossRef](#)]
30. Choi, Y.; Gehring, M.; Johnson, L.; Hannon, M.; Harada, J.J.; Goldberg, R.B.; Jacobsen, S.E.; Fischer, R.L.; California, B. DEMETER, a DNA Glycosylase Domain Protein, Is Required for Endosperm Gene Imprinting and Seed Viability in Arabidopsis Replicates to form a Syncytium, and Following Cellularization, Produces storage Proteins, Lipids, and Starch. *Cell* **2002**, *110*, 33–42. [[CrossRef](#)]
31. Ibarra, C.A.; Feng, X.; Schoft, V.K.; Hsieh, T.F.; Uzawa, R.; Rodrigues, J.A.; Zemach, A.; Chumak, N.; Machlicova, A.; Nishimura, T.; et al. Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science (80)* **2012**, *337*, 1360–1364. [[CrossRef](#)]
32. Li, Y.; Kumar, S.; Qian, W. Active DNA demethylation: Mechanism and role in plant development. *Plant Cell Rep.* **2018**, *37*, 77–85. [[CrossRef](#)] [[PubMed](#)]
33. Slotkin, R.K.; Vaughn, M.; Borges, F.; Tanurdzić, M.; Becker, J.D.; Feijó, J.A.; Martienssen, R.A. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **2009**, *136*, 461–472. [[CrossRef](#)] [[PubMed](#)]
34. Walker, J.; Gao, H.; Zhang, J.; Aldridge, B.; Vickers, M.; Higgins, J.D.; Feng, X. Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. *Nat. Genet.* **2018**, *50*, 130–137. [[CrossRef](#)]
35. Olmedo-Monfil, V.; Durán-Figueroa, N.; Arteaga-Vandázquez, M.; Demesa-Arévalo, E.; Autran, D.; Grimanelli, D.; Slotkin, K.; Martienssen, R.A.; Vielle-Calzada, J.-P. Control of female gamete formation by a small RNA pathway in Arabidopsis HHS Public Access. *Nature* **2010**, *464*, 628–632. [[CrossRef](#)]
36. Han, Y.Z.; Huang, B.Q.; Zee, S.Y.; Yuan, M. Symplastic communication between the central cell and the egg apparatus cells in the embryo sac of *Torenia fournieri* Lind. before and during fertilization. *Planta* **2000**, *211*, 158–162. [[CrossRef](#)]
37. Saze, H.; Scheid, O.M.; Paszkowski, J. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat. Genet.* **2003**, *34*, 65–69. [[CrossRef](#)]
38. Jablonka, E.; Raz, G. Transgenerational Epigenetic Inheritance: Prevalence, Mechanisms, and Implications for the Study of Heredity and Evolution. *Q. Rev. Biol.* **2009**, *84*, 131–176. [[CrossRef](#)]
39. Feng, S.; Jacobsen, S.E.; Reik, W. Epigenetic reprogramming in plant and animal development. *Science (80-)* **2010**, *330*, 622–627. [[CrossRef](#)]
40. Herrera, C.M.; Medrano, M.; Bazaga, P. Comparative spatial genetics and epigenetics of plant populations: Heuristic value and a proof of concept. *Mol. Ecol.* **2016**, *25*, 1653–1664. [[CrossRef](#)]
41. Niederhuth, C.E.; Bewick, A.J.; Ji, L.; Alabady, M.; Kim, K.D.; Page, J.T.; Li, Q.; Rohr, N.A.; Rambani, A.; Burke, J.M.; et al. Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* **2016**, *17*, 194. [[CrossRef](#)]
42. Wilschut, R.A.; Oplaat, C.; Snoek, L.B.; Kirschner, J.; Verhoeven, K.J.F. Natural epigenetic variation contributes to heritable flowering divergence in a widespread asexual dandelion lineage. *Mol. Ecol.* **2016**, *25*, 1759–1768. [[CrossRef](#)] [[PubMed](#)]
43. Herman, J.J.; Sultan, S.E. Adaptive transgenerational plasticity in plants: Case studies, mechanisms, and implications for natural populations. *Front. Plant Sci.* **2011**, *2*, 102. [[CrossRef](#)] [[PubMed](#)]
44. Herrera, C.M.; Bazaga, P. Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and nonprickly leaves in heterophyllous *Ilex aquifolium* (Aquifoliaceae) trees. *Bot. J. Linn. Soc.* **2013**, *171*, 441–452. [[CrossRef](#)]
45. Medrano, M.; Herrera, C.M.; Bazaga, P. Epigenetic variation predicts regional and local intraspecific functional diversity in a perennial herb. *Mol. Ecol.* **2014**, *23*, 4926–4938. [[CrossRef](#)]
46. Colicchio, J.M.; Miura, F.; Kelly, J.K.; Ito, T.; Hileman, L.C. DNA methylation and gene expression in *Mimulus guttatus*. *BMC Genom.* **2015**, *16*, 507. [[CrossRef](#)]
47. Agrawal, A.A.; Strauss, S.Y.; Stout, M.J. Costs of induced responses and tolerance to herbivore in male and female fitness components of wild radish. *Evolution (N. Y.)* **1999**, *53*, 1093–1104.

48. Bilichak, A.; Kovalchuk, I. Transgenerational response to stress in plants and its application for breeding. *J. Exp. Bot.* **2016**, *67*, 2081–2092. [CrossRef]
49. Ramírez-Carrasco, G.; Martínez-Aguilar, K.; Alvarez-Venegas, R. Transgenerational defense priming for crop protection against plant pathogens: A hypothesis. *Front. Plant Sci.* **2017**, *8*, 696. [CrossRef]
50. Latzel, V.; Rendina González, A.P.; Rosenthal, J. Epigenetic memory as a basis for intelligent behavior in clonal plants. *Front. Plant Sci.* **2016**, *7*, 1–7. [CrossRef]
51. Dewan, S.; De Frenne, P.; Broeck, A.V.; Steenackers, M.; Mijnsbrugge, K.V.; Verheyen, K. Transgenerational effects in asexually reproduced offspring of *Populus*. *PLoS ONE* **2018**, *13*, 1–16. [CrossRef]
52. FAOSTAT. 2017. FAOSTAT. In FAOSTAT. Available online: <http://www.fao.org/faostat/fr/#data/QC> (accessed on 4 May 2019).
53. Lane, W.D. *Micropropagation of Apple (Malus Domestica Barkh.)*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 229–243.
54. Visser, T. *Juvenile Phase and Growth of Apple and Pear Seedlings*; Springer: Berlin, Germany, 1964; Volume 13.
55. Lavee, S.; Avidan, N.; Haskal, A.; Ogrodovich, A. Juvenility Period Reduction in Olive Seedlings—A Tool for Enhancement of Breeding. *Olivae* **1996**, *60*, 33–41.
56. Hackett, W.P.; Murray, J.R. *Maturation and Rejuvenation in Woody Plants*; Springer: Berlin, Germany, 2015; ISBN 0870554921.
57. Rugini, E. Olive (*Olea eu-ropaea* L.). In *Biotechnology in Agri-Culture and Forestry*; Springer: Berlin, Germany, 1986; pp. 253–267.
58. Bellini, E. Variabilidad genetica heredabilidad de algunos caracteres en plantas de semillas de olive obtenidas por cruzamiento. *Olivae Rev. Of. del Cons. Oleícola Int.* **1993**, *49*, 21–34.
59. Meilan, R. *Floral Induction in Woody Angiosperms*; Kluwer Academic Publishers: Berlin, Germany, 1997; Volume 14.
60. Telfer, A.; Bollman, K.M.; Poethig, R.S. Phase change and the regulation of trichome distribution in *Arabidopsis thaliana*. *Development* **1997**, *124*, 645–654. [PubMed]
61. Poethig, R.S. Phase Change and the Regulation of Developmental Timing in Plants. *Science (80)* **2003**, *301*, 334–336. [CrossRef] [PubMed]
62. Wang, J.-W.; Park, M.Y.; Wang, L.-J.; Koo, Y.; Chen, X.-Y.; Weigel, D.; Poethig, R.S. MiRNA Control of Vegetative Phase Change in Trees. *PLoS Genet.* **2011**, *7*, e1002012. [CrossRef]
63. Critchfield, W.B. Leaf dimorphism in *Populus trichocarpa*. *Am. J. Bot.* **1960**, *47*, 699–711. [CrossRef]
64. Lohse, M.; Nagel, A.; Herter, T.; May, P.; Schroda, M.; Zrenner, R.; Tohge, T.; Fernie, A.R.; Stitt, M.; Usadel, B. Mercator: A fast and simple web server for genome scale functional annotation of plant sequence data. *Plant Cell Environ.* **2014**, *37*, 1250–1258. [CrossRef]
65. Daccord, N.; Celton, J.-M.; Linsmith, G.; Becker, C.; Choisine, N.; Schijlen, E.; van de Geest, H.; Bianco, L.; Micheletti, D.; Velasco, R.; et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* **2017**, *49*, 1099–1106. [CrossRef]
66. Basheer-Salimia, R. Juvenility, Maturation, and Rejuvenation in Woody Plants. *Hortic. Rev. (Am. Soc. Hortic. Sci.)* **2007**, 109–155.
67. Zimmerman, R.H. Juvenility and flowering of fruit trees. In Proceedings of the Acta Horticulturae 34: Symposium on Growth Regulators in Fruit Production, Beltsville, MD, USA, 1 December 1973; pp. 139–142.
68. Hanke, M.-V.; Flachowsky, H.; Peil, A.; Hättasch, C. No flower no fruit—Genetic potentials to trigger flowering in fruit trees. *Genes Genomes Genom.* **2007**, *1*, 1–20.
69. Zhang, X.Z.; Zhao, Y.B.; Li, C.M.; Chen, D.M.; Wang, G.P.; Chang, R.F.; Shu, H.R. Potential polyphenol markers of phase change in apple (*Malus domestica*). *J. Plant Physiol.* **2007**, *164*, 574–580. [CrossRef] [PubMed]
70. Murray, J.R.; Smith, A.G.; Hackett, W.P. Differential dihydroflavonol reductase transcription and anthocyanin pigmentation in the juvenile and mature phases of ivy (*Hedera helix* L.). *Planta* **1994**, *194*, 102–109. [CrossRef]
71. Hand, P.; Besford, R.T.; Richardson, C.M.; Peppitt, S.D. Antibodies to phase related proteins in juvenile and mature *Prunus avium*. *Plant Growth Regul.* **1996**, *20*, 25–29. [CrossRef]
72. Ryan, M.G.; Binkley, D.; Fownes, J.H. Age related decline in forest productivity: Pattern and process. *Adv. Ecol. Res.* **1997**, *27*, 213–262.
73. Day, M.E.; Greenwood, M.S.; Diaz-Sala, C. Age- and size-related trends in woody plant shoot development: Regulatory pathways and evidence for genetic control. *Tree Physiol.* **2002**, *22*, 507–513. [CrossRef] [PubMed]

74. Bond, B.J. Age-related changes in photosynthesis of woody plants. *Trends Plant Sci.* **2000**, *5*, 349–353. [CrossRef]
75. Greenwood, M.S. Juvenility and maturation in conifers. *Curr. Concepts* **1995**, *15*, 433–438.
76. Greenwood, M.S. Phase change in loblolly pine: Shoot development as a function of age. *Physiol. Plant.* **1984**, *61*, 518–522. [CrossRef]
77. Hutchison, K.W.; Sherman, C.D.; Weber, J.; Smith, S.S.; Singer, P.B.; Greenwood, M.S. Maturation in Larch: II. Effects of Age on Photosynthesis and Gene Expression in Developing Foliage. *Plant Physiol.* **1990**, *94*, 1308–1315. [CrossRef]
78. Rebbeck, J.; Jensen, K.F.; Greenwood, M.S. Ozone effects on grafted mature and juvenile red spruce: Photosynthesis, stomatal conductance, and chlorophyll concentration. *Can. J. For. Res.* **1993**, *23*, 450–456. [CrossRef]
79. McGowran, E.; Douglas, G.C.; Parkinson, M. Morphological and physiological markers of juvenility and maturity in shoot cultures of oak (*Quercus robur* and *Q. petraea*). *Tree Physiol.* **1998**, *18*, 251–257. [CrossRef] [PubMed]
80. Zhang, X.; Yazaki, J.; Sundaresan, A.; Cokus, S.; Chan, S.W.L.; Chen, H.; Henderson, I.R.; Shinn, P.; Pellegrini, M.; Jacobsen, S.E.; et al. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in *Arabidopsis*. *Cell* **2006**, *126*, 1189–1201. [CrossRef] [PubMed]
81. Zilberman, D.; Gehring, M.; Tran, R.K.; Ballinger, T.; Henikoff, S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* **2007**, *39*, 61–69. [CrossRef] [PubMed]
82. Lespinasse, Y.; Bouvier, L.; Djulbic, M.; Chevreau, E. Haploidy in apple and pear. *Acta Hortic.* **1998**, *484*, 49–54. [CrossRef]
83. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [CrossRef]
84. Xi, Y.; Li, W. BSMAP: Whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **2009**, *10*, 232. [CrossRef]
85. R Core Team R. *A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.
86. Buels, R.; Yao, E.; Diesh, C.M.; Hayes, R.D.; Munoz-Torres, M.; Helt, G.; Goodstein, D.M.; Elsik, C.G.; Lewis, S.E.; Stein, L.; et al. JBrowse: A dynamic web platform for genome visualization and analysis. *Genome Biol.* **2016**, *17*, 66. [CrossRef]
87. Depuydt, S.; Trenkamp, S.; Fernie, A.R.; Elftieh, S.; Renou, J.-P.; Vuylsteke, M.; Holsters, M.; Vereecke, D. An Integrated Genomics Approach to Define Niche Establishment by *Rhodococcus fascians*. *Plant Physiol.* **2009**, *149*, 1366–1386. [CrossRef]
88. Smyth, G.K.; Michaud, J.; Scott, H.S. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **2005**, *21*, 2067–2075. [CrossRef]
89. Untergasser, A.; Nijveen, H.; Rao, X.; Bisseling, T.; Geurts, R.; Leunissen, J.A.M. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids* **2007**, *35*, W71–W74. [CrossRef]
90. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **2001**, *29*, e45. [CrossRef] [PubMed]
91. Vandesompele, J.; De Preter, K.; Pattyn, F.; Poppe, B.; Van Roy, N.; De Paepe, A.; Speleman, F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **2002**, *3*, RESEARCH0034. [CrossRef] [PubMed]
92. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc.* **1995**, *57*, 289–300. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

3.3. Résumé de l'article

Dans l'article présenté précédemment nous avons comparé la transmission des marques épigénétiques dans le cas d'une reproduction sexuée et d'une propagation asexuée chez le pommier (*Malus domestica*). Nous avons également identifié un effet potentiel d'une différence de méthylation sur la transcription des gènes.

Dans un premier temps nous avons observé des différences phénotypiques que nous avons pu relier au stade de développement des plantes. Les plants issus de semis (SD) présentaient un état juvénile identifiable par une très faible densité de trichome sur la face inférieure des feuilles. Dans le cas des plants greffés (YG) et adultes (OG), nous avons identifié une densité de trichomes plus importante sur la face inférieure de leur feuille, caractère que nous avons pu relier à un état mature des plantes d'après la littérature. Ces observations nous ont permis de confirmer que le greffage n'est pas comparable à un processus de rajeunissement complet, comme ce qui est observé dans le cadre d'une multiplication sexuée, avec l'obtention de semis.

En parallèle de cette étude phénotypique, nous avons également réalisé une analyse du transcriptome des plantes. Nous avons ainsi pu constater que le niveau de transcription des gènes, lié à la photosynthèse, était plus élevé dans les plants issus de semis (SD) par rapport à l'arbre adulte (OG), tandis que les plants nouvellement greffés (YG) affichaient un niveau de transcription intermédiaire.

Enfin, nous avons analysé les données de méthylation issues du séquençage bisulfite de chacun de nos échantillons. Nous avons ainsi pu montrer dans un premier temps que le niveau de méthylation d'ADN global dans les plants issus de semis ou les plants greffés n'est pas modifié par la multiplication sexuée ou par la propagation asexuée par rapport à l'arbre adulte. Nous avons cependant identifié des modifications locales, que nous avons appelé DMRs (région différentiellement méthylées). Ces différences ont été retrouvées en majorité dans le contexte de méthylation CHH. Il semble donc que le processus de reprogrammation de la méthylation ayant lieu pendant la méiose n'affecte pas le niveau global de méthylation du génome, assurant ainsi la stabilité du génome notamment par la répression des éléments transposables. Il semble, cependant qu'il existe un processus de réarrangement de la méthylation sur des régions spécifiques lors de la multiplication sexuée. Ces réarrangements pourraient être impliqués dans une augmentation de la compétitivité des plants issus de semis, leur permettant ainsi un développement plus rapide dans un milieu donné par rapport à un arbre adulte. Cette observation était particulièrement marquante pour les gènes associés à la photosynthèse.

L'analyse des données transcriptomiques et des données de méthylation de l'ADN, a également permis de montrer que les plantes greffées (YG) étaient à l'interphase entre l'arbre (OG) et les semis (SD).

Les résultats que nous avons publiés dans cet article tendent à montrer que, d'un point de vue physiologique, transcriptomique et épigénétique, les plantes nouvellement greffées (YG) sont à l'interphase entre un arbre (OG) et un semis (SD), présentant des caractéristiques qui sont particulières à la fois aux stades immatures et matures de la plante.

4. Partie 2 - Résultats complémentaires

La vie est affaire de choix et la science n'échappe pas à cette règle, c'est pourquoi nous n'avons sélectionné qu'une partie de nos résultats pour la publication présentée précédemment. Nous avons retenu les résultats qui nous semblaient le plus pertinent et le plus marquant, comme la densité des trichomes pour le phénotypage. Mais j'ai généré d'autres données, de phénotypage et également travaillé sur plusieurs façons de calculer des régions différentiellement méthylées (DMRs) que je vais détailler par la suite.

4.1. Matériels et méthodes

4.1.1. Matériel végétal

Pour cette partie nous avons utilisé le même matériel végétal que ce que nous avons présenté précédemment à savoir : OG, YG et SD. Des mesures et prélèvements supplémentaires sur ces échantillons ont été réalisés pour les expériences de phénotypage complémentaire.

4.1.2. Fixation et coupes histologiques des échantillons YG et SD.

4 feuilles de 4 plants issus de semis et de 4 plants issus de greffes ont été prélevées. Sur chaque feuille deux prélèvements sont effectués : Le premier dans le limbe des feuilles, le second au niveau de la nervure centrale comme présenté dans la Figure 4.1-1. Chaque fragment est ensuite fixé en glutaraldéhyde puis inclus en résine Technovit7100 en suivant le même protocole d'inclusion que (Celton *et al.*, 2014). Brièvement, les échantillons sont incubés sous vide dans un tampon de glutaraldéhyde à 4°C pendant 3h. Une fois fixé, les échantillons sont déshydratés dans des bains d'éthanol successifs puis inclus en résine Technovit7100 (Kulzer Histo-technique kit; Labonord). Des coupes de 3µm d'épaisseur sont réalisées à l'aide d'un microtome et montées sur des lames en verre de microscope. Une coloration au bleu de toluidine est ensuite réalisée avant de déshydrater les échantillons et de monter une lamelle. Des observations sont faites sous microscope optique Leica DM1000. Des acquisitions d'image sont réalisées à l'aide du logiciel d'histologie QCapturePro®. Les photos sont ensuite analysées à l'aide du logiciel ImageJ® pour déterminer le diamètre de la nervure centrale et l'épaisseur du limbe.

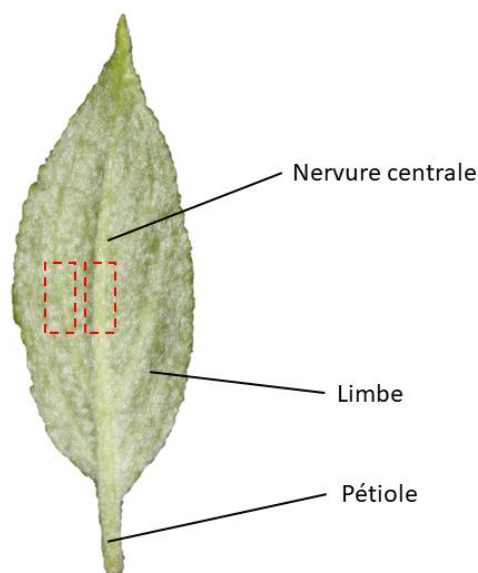


Figure 4.1-1: illustration du prélèvement sur une feuille de pommier. Les tissus prélevés pour être fixé sont indiqués par un rectangle rouge.

4.1.3. Microarray et calcul des régions différenciellement méthylées

Les extractions ADNs et ARNs ont été réalisées et décrites dans la publication précédente (Perrin *et al.*, 2020). Les analyses de microarray et de calculs des régions différenciellement méthylées réalisées dans cette partie, ont été effectuées en suivant les mêmes paramètres que ceux présentés dans la publication précédente (Perrin et al 2020).

4.2. Résultats

4.2.1. Analyse de la troisième comparaison : plants greffés versus plants semés

Dans la publication précédemment présentée, nous avons discuté de la transmission verticale des marques épigénétiques de l'arbre adulte vers les plants greffés et les plants issus de semis. Nous avons cherché à comprendre comment les marques épigénétiques sont transmises via une multiplication sexuée et asexuée.

Dans cette partie je vais présenter une troisième comparaison : les plants greffés (YG) par rapport aux plants semés (SD). Pour cela j'ai étudié différents caractères phénotypiques, puis je me suis intéressé aux différences de niveaux transcriptionnels et épigénétiques des plants semés par rapport aux plants greffés. Avec l'étude de cette comparaison nous espérons retrouver des résultats similaires à ceux que nous avons précédemment obtenus (Perrin *et al.*, 2020). Si tel est le cas, des études futures, où l'accès aux plants parents n'est pas possible (perte du parent, impossibilité de prélèvement, etc...) pourront tout de même être réalisées sur des semis et des greffons en extrapolant les résultats pour les plants parents.

a) Phénotypage

Pendant ma thèse j'ai, en premier lieu cherché des moyens macroscopiques et microscopiques permettant de caractériser les jeunes plants semés et les greffons. Pour cela j'ai effectué le phénotypage de plants ayant poussés en serre pour contrôler l'effet environnemental.

Pour phénotyper des plants semblables (taille similaire) les semis ont été réalisés trois semaines avant le greffage. Ceci permet d'obtenir au bout de 2 ou 3 mois de croissance des plants issus de semis et des greffons présentant des tailles de tige principale similaires lors des prélèvements.

Phénotypage macroscopique

Dans un premier temps je me suis attaché à montrer les différences de phénotype visibles à l'œil nu. J'ai commencé par compter le nombre d'étages foliaires sur des greffons et des plants issus de semis. Pour ce faire j'ai décapité neuf plants de chaque groupe et je les ai ensuite laissés pousser pendant trois semaines avant de compter le nombre d'étages foliaires à partir du point de décapitation. J'ai effectué un second comptage deux semaines plus tard. Je n'ai pas observé de différences significatives du nombre d'étages foliaires au bout de trois semaines (Figure 4.2-1A), mais j'ai observé une différence significative (à 5%) du nombre d'étages foliaires au bout de cinq semaines (Figure 4.2-1B). Les jeunes plants présentaient alors en moyenne vingt étages foliaires, contre moins de dix-huit chez les greffons.

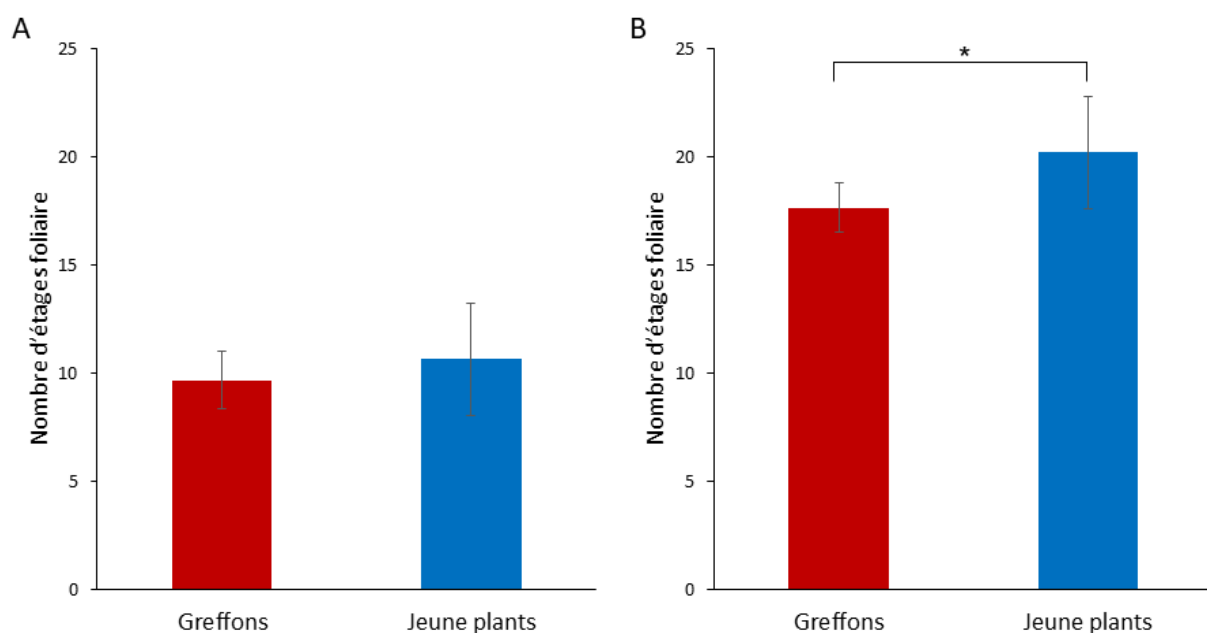


Figure 4.2-1 : comptage du nombre d'étages foliaires. Nombre d'étages foliaires chez les greffons et le jeune plant issu de semis. Au bout de trois (A) ou cinq (B) semaines de croissance après décapitation. Un test de student a été réalisé pour évaluer la différence entre les groupes ou chaque date. *Astérisque p-value*: *: 5% n=9.

Par la suite j'ai prélevé des feuilles de trois âges différents sur les plants issus de semis et des greffons (Figure 4.2-2). Je n'ai pas pu effectuer de prélèvements sur l'arbre adulte de référence car il subissait à ce moment une très forte attaque d'acariens rouges (*Panonychus ulmi*) et de thrips (*Frankliniella occidentalis*) rendant le prélèvement de feuilles saines impossible.

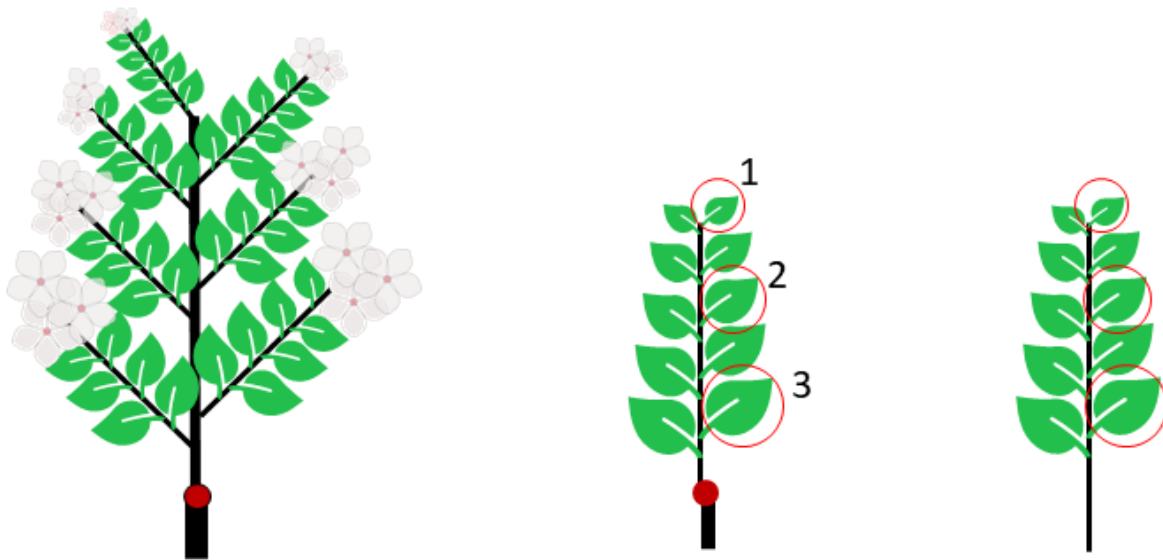


Figure 4.2-2: plan de prélèvement : cercle rouge: feuilles prélevées, les chiffres indiquent les trois stades auquel les feuilles ont été prélevées. Stade 1: dernière feuille complètement formée/déroulée; stade 2: feuille d'âge moyen, présentant un parenchyme de couleur vert clair similaire aux feuilles jeunes; stade 3: feuille âgée plus rigide, présentant un parenchyme de couleur vert foncé

Sur les prélèvements réalisés (Figure 4.2-3) on peut observer une différence très nette entre la forme des feuilles issues de plants semés (Figure 4.2-3B, D et F) et de plants greffés (Figure 4.2-3A, C et D). Les feuilles des plants semés ont des formes plus lancéolées que celles issues de plants greffés. On remarque également que les feuilles d'âge moyen à vieux (prélèvements numéro 2 et 3 sur Figure 4.2-2) issues de plants greffés (Figure 4.2-3C et E) ont des bordures ou serrations moins marquées que les feuilles issues de plants semés. La serration correspond à l'aspect dentelé de la bordure des feuilles.

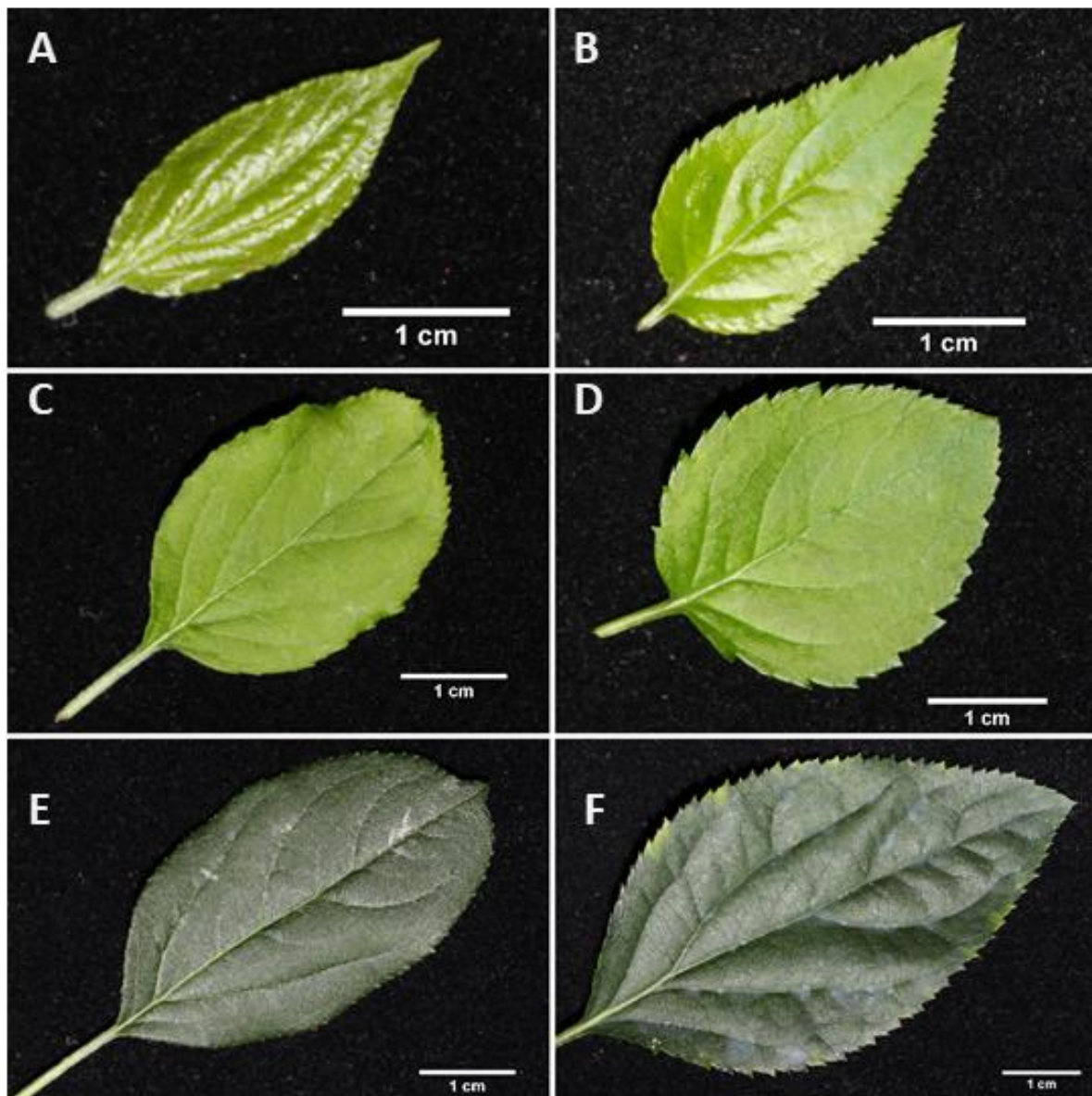


Figure 4.2-3: photos macroscopiques de feuilles de pommier : Panel de feuilles issues de greffon (A, C, E) et de semis (B, D, F). Trois âges différents de feuilles sont présentés ici, correspondant aux stades précédemment définis (Figure 4.2-2): stade 1 (A et B), stade 2 (C et D) et stade 3 (E et F).

A l'aide d'une loupe binoculaire, on observe que les serrations des feuilles issues de plants greffés sont présentes, mais elles sont moins profondes et moins ouvertes (Figure 4.2-4).

En observant les faces inférieures et supérieures des feuilles plus âgées (Figure 4.2-4 C à F) on observe que le réseau de veines est plus dense et plus serré sur les feuilles issues de plants greffés que sur les feuilles issues de plants semés.

Lors du phénotypage à l'échelle macroscopique des jeunes plants semés et greffés, j'ai pu mettre en évidence des différences phénotypiques marquées. J'ai ensuite voulu savoir si l'on pouvait retrouver des différences phénotypiques à l'échelle microscopique.

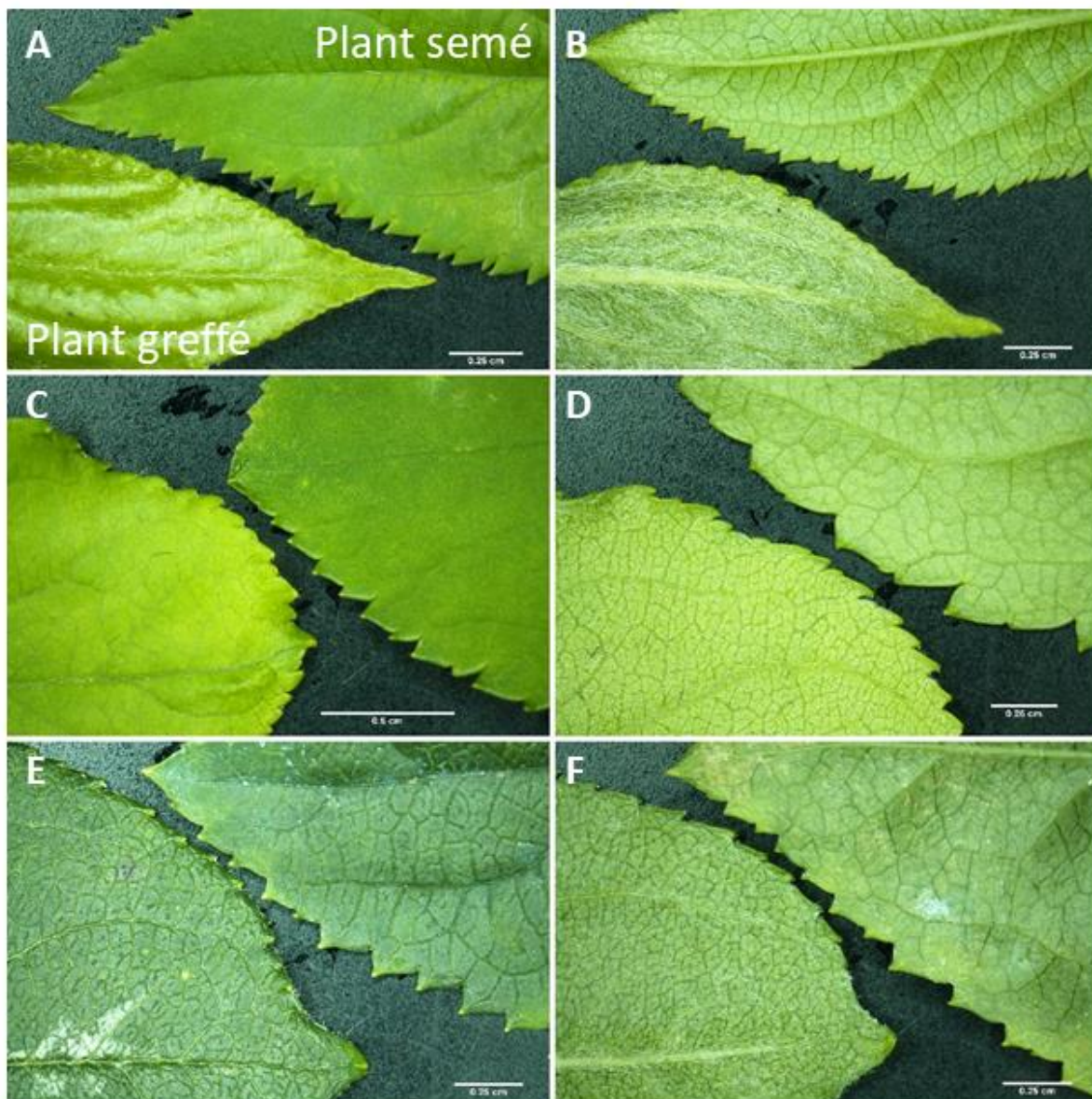


Figure 4.2-4: photos de feuilles de pommier réalisées à la loupe binoculaire . . Les photos A, C et E montrent la face adaxiale des feuilles, les photos B, D et F montrent la face abaxiale des feuilles. Sur chaque photo, la feuille dans le coin en bas à gauche est issue des greffons et la feuille dans le coin en haut à droite est une feuille issue de plants semés comme indiqué en (A). Trois âges différents de feuilles sont présentés ici, correspondant aux stades précédemment définis (Fig. 1): stade 1 (A et B), stade 2 (C et D) et stade 3 (E et F).

Phénotypage microscopique

Afin d'identifier des différences sur les structures internes de la feuille, j'ai réalisé des coupes histologiques de jeunes feuilles. J'ai observé deux types de tissus, du limbe et la nervure centrale des feuilles échantillonnées. On peut observer des différences morphologiques entre les coupes de feuilles de plants greffés et celles de plants semés. La nervure centrale des plants semés (Figure 4.2-5A) est beaucoup plus arrondie que celle des plants issus de greffes (Figure 4.2-5B). Les tissus vasculaires xylème et phloème semblent également plus développés chez le greffon que chez les plants semés. J'ai effectué des colorations sur ces tissus avec du rouge carmin (coloration du xylème) et du bleu de méthylène (coloration du phloème), mais les tests ont été non concluant, je n'ai pas pu observer mes échantillons avec ces colorations. Nous pouvons cependant estimer cette différence via la comparaison du diamètre entre les échantillons (Figure 4.2-6A). Les plants issus de semis présentent un diamètre de la nervure centrale inférieur de 144 μm en moyenne par rapport aux greffons (différence significative à 1% suite à un test de Student).

Dans un second temps j'ai observé la morphologie du limbe (Figure 4.2-5C & D). Je n'ai pas observé de différence apparente. J'ai également réalisé une mesure de l'épaisseur du limbe qui laisse apparaître une petite différence (Figure 4.2-6B) de 4 μm en moyenne entre les deux échantillons (différence significative à 5% suite à un test de Student).

Nous avons donc identifié plusieurs différences phénotypiques entre plants greffés et plants issus de semis. A partir de ces constatations nous avons cherché à identifier les mécanismes moléculaires sous-jacent ; comme l'effet de la méthylation de l'ADN ou de la différence d'expression des gènes.

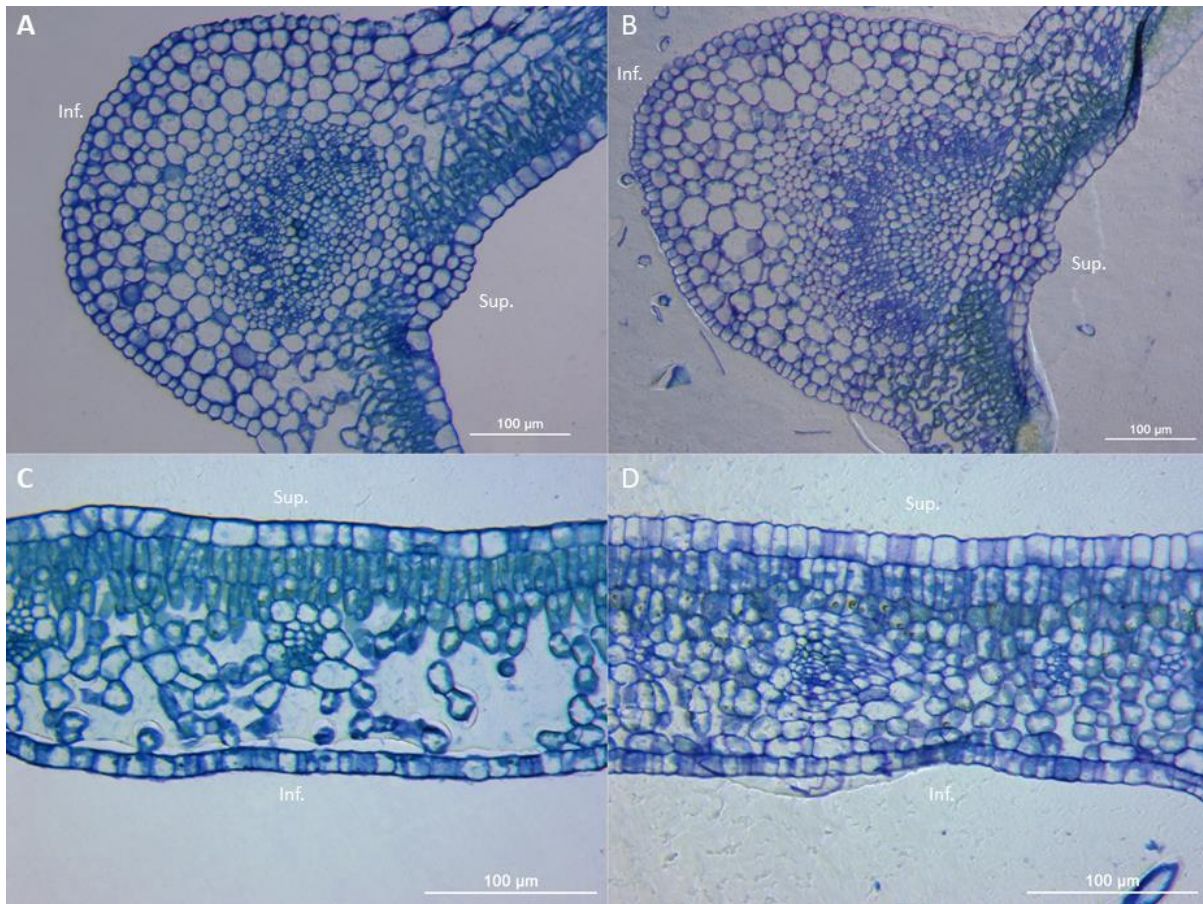


Figure 4.2-5: Exemples de coupes transversales de jeunes feuilles (stade 1, Fig. 1) issues de plants de semis (A et C) et de plants greffés (B et D). En (A) et (B) coupe transversale de la veine centrale et en (C) et (D) coupe transversale du limbe. Inf. : face inférieure (abaxiale) de la feuille; Sup. : face supérieure (adaxiale) de la feuille. Coupe de 3 μ m colorée au bleu de toluidine.

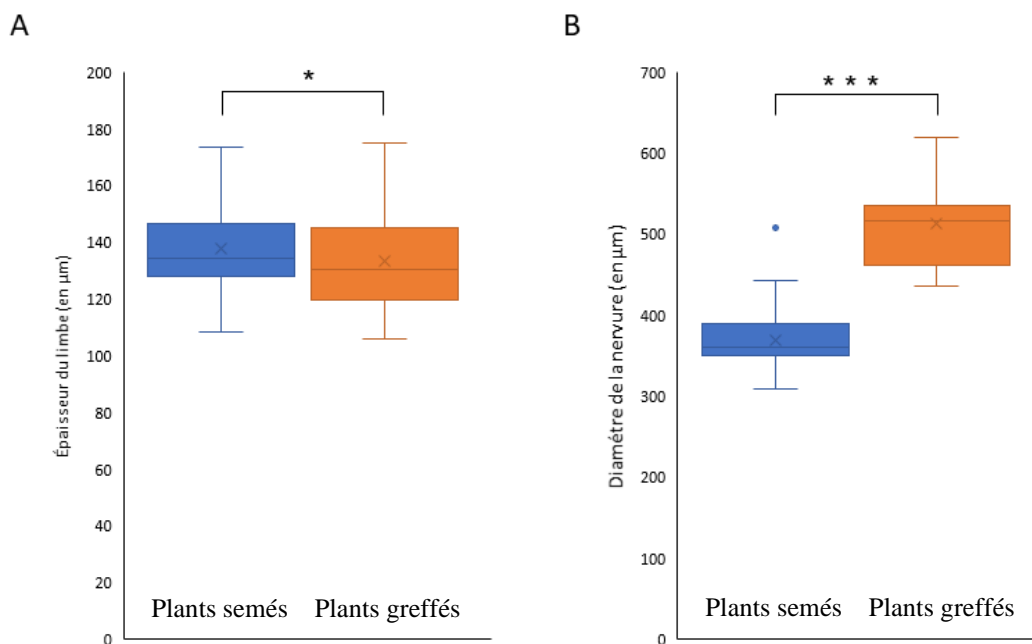


Figure 4.2-6: mesure de l'épaisseur du limbe (A.) et du diamètre de la nervure centrale (B.). Un test de student a été réalisé pour évaluer la différence entre les groupes à chaque date de prélèvement. Astérisque *p*-value: *: 5%, ** 1%, ***1‰ n=200.

b) Transcrits différentiellement exprimés

Afin de mieux comprendre l'impact du mode de reproduction sur l'expression des gènes, nous avons réalisé une analyse transcriptomique pour cette comparaison afin d'analyser les transcrits différentiellement exprimés (DETs).

Pour cela nous avons utilisé une puce micro-array basée sur la nouvelle annotation du pommier.

Pour la suite des analyses des DETs dans ce manuscrit, nous avons exclu les séquences en antisens ce qui induit une différence du nombre de DETs par rapport à la publication précédente, afin de simplifier l'analyse.

Nous avons identifié 5396 DETs associés à des gènes (DEG) dans la comparaison YGvsSD (Table 4.2-1), soit respectivement 9,8 et 18,9% de DEGs en plus que dans les comparaisons SDvsOG et OGvsYG.

Concernant les DETs associés des éléments transposables (DETE) nous en avons identifié 670 dans la comparaison YGvsSD, soit 27% de DETEs en plus que dans la comparaison SDvsOG et 10% de moins que dans la comparaison OGvsYG.

Table 4.2-1 : Nombre de DEGs et DETEs par comparaison et en communs entre les comparaisons. L'échantillon utilisé comme référence pour le calcul est souligné dans chaque comparaison. Les nombres indiqués entre parenthèses correspondent au DET plus/moins transcrit par rapport à la référence, respectivement.

	<i>DEG</i>	<i>DETE</i>
DET spécifique à chaque comparaison		
<u>YGvsSD</u>	2.606 (1.618/988)	454
<u>SDvsOG</u>	1.353 (329/258)	232
OGvsYG	1.958 (410/177)	508
DET communs		
YGvs <u>SD</u> – <u>SD</u> vsOG	1.864 (1.588/276)	114
<u>YGvsSD</u> – <u>OGvsYG</u>	339 (273/54)	89
OGvsYG – <u>SDvsOG</u>	1.062 (271/789)	128
YGvsSD – <u>SDvsOG</u> – OGvsYG	587	13

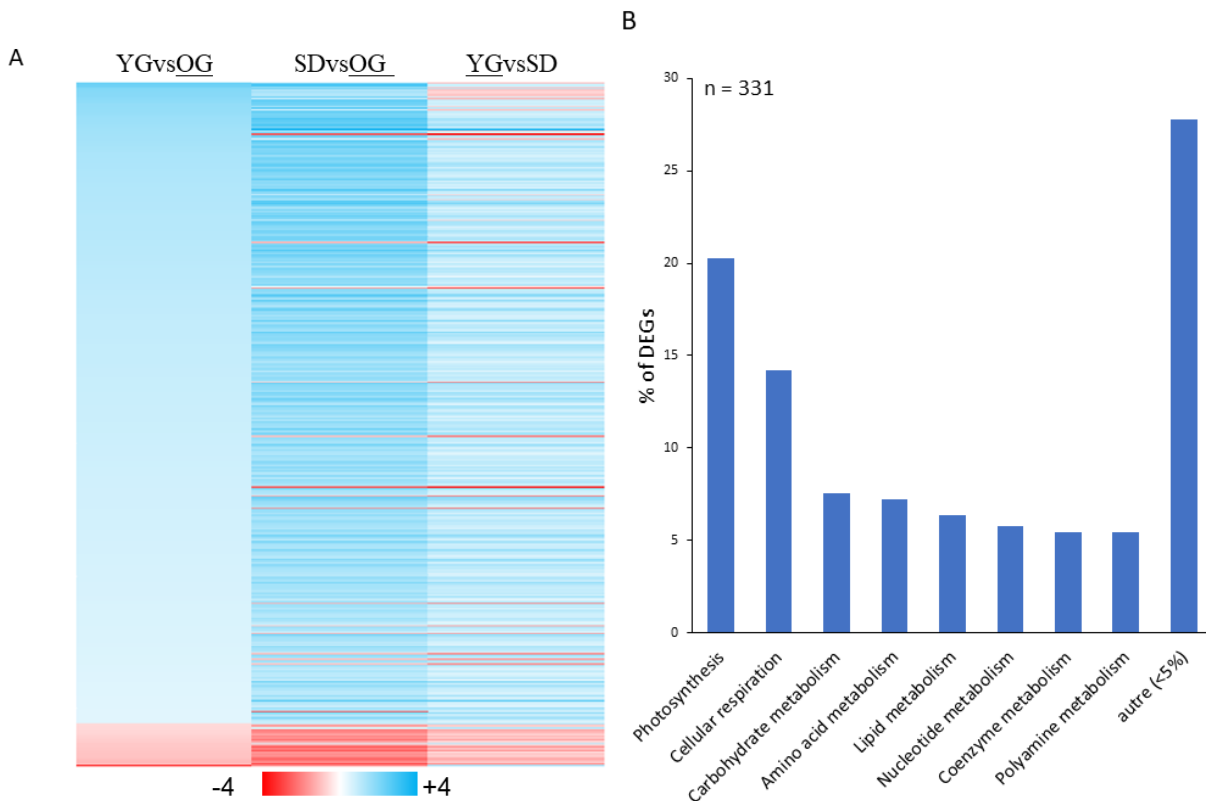


Figure 4.2-7 : DEGs communs aux trois comparaisons. A. Ratio d'expression des DEGs communs entre les trois comparaisons (n=587). L'échantillon de référence est souligné. B. Classification des DEGs communs aux trois comparaisons par code BIN. Les catégories inférieures à 5% sont dans la catégorie autre. Les DEGs de la catégorie 35 (non classifié) ne sont pas présentés.

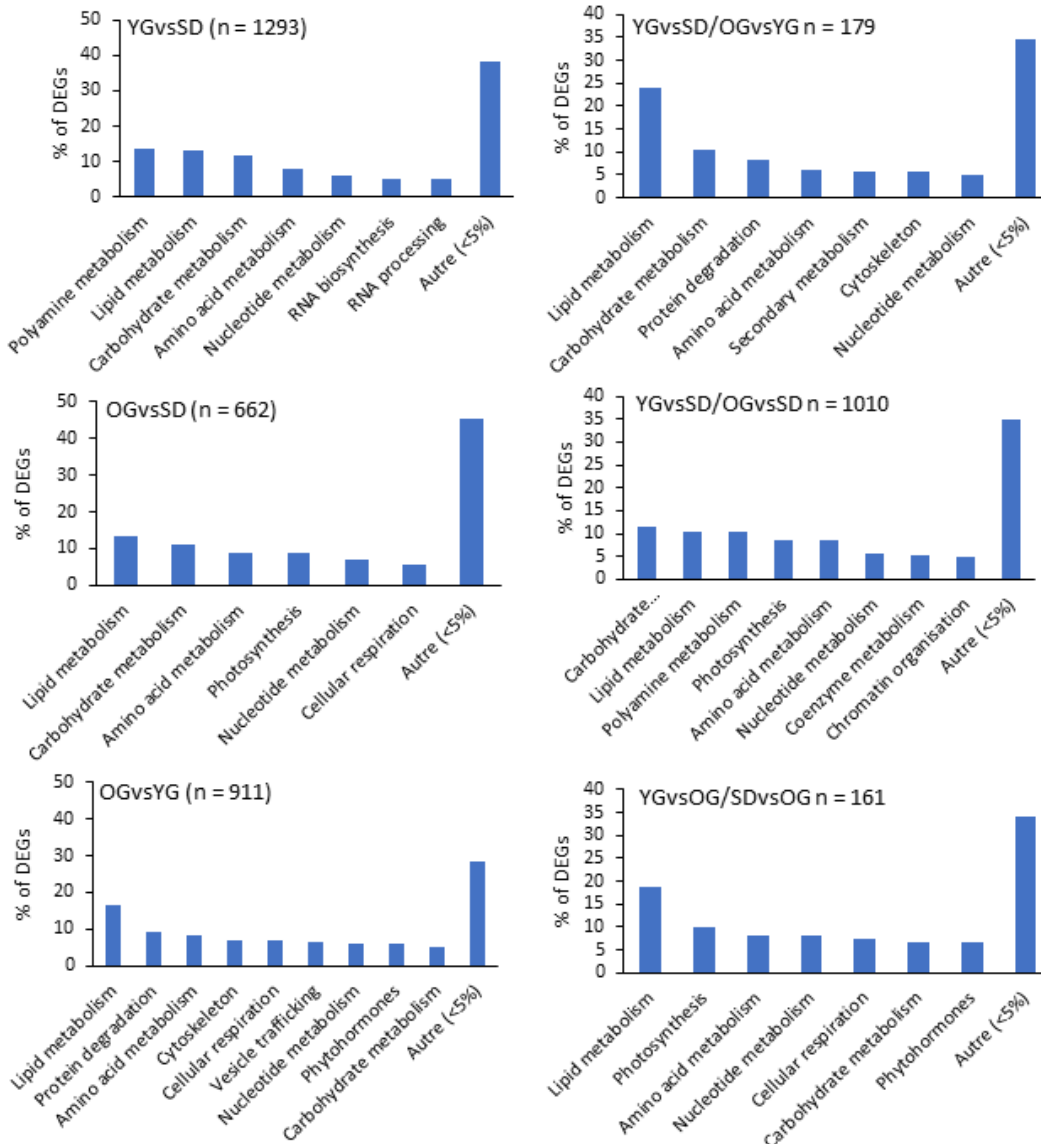


Figure 4.2-8 : classification BIN des DEGs spécifiques à chaque comparaison et en commun entre deux comparaisons. A. Classification des DEGs spécifiques à chaque comparaison en fonction de leur code BIN. B. Classification des DEGs communs entre deux comparaisons. n: le nombre de DEGs dans chaque graphique. Les catégories représentant moins de 5% sont dans la catégorie autre. Les DEGs de la catégorie 35 (non classifié) ne sont pas présentées par souci de lisibilité.

On observe également 1864 DEGs en communs entre les comparaisons YGvsSD et SDvsOG. Il y a plus de DEGs en communs entre ces comparaisons qu'entre les comparaisons YGvsSD et OGvsYG (339 DEGs) et OGvsYG et SDvsOG (1.062 DEGs) réunies. Ce qui suggère que le greffon est plus proche de l'arbre adulte mature en termes de niveau d'expression que le semis.

Je constate que lorsque l'on considère deux comparaisons, un DEG est quasiment soit sur-exprimé soit sous-exprimé pour l'échantillon en commun dans les deux comparaisons (Figure 4.2-7A) Seul 12 DEGs en communs entre SDvsOG et OGvsYG et 2 DEGs pour OGvsYG/SDvsOG ne suivent pas cette règle.

Je me suis ensuite intéressé aux DEGs communs entre les trois comparaisons (Figure 4.2-7A). J'observe que la majorité des DEGs sont sur-exprimés dans les semis et les jeunes greffons par rapport à l'arbre adulte dans les comparaisons SDvsOG et OGvsYG. De la même façon pour SDvsOG et YGvsSD j'observe que la majorité des DEGs sont sur-exprimés dans les plants semés.

Enfin pour les comparaisons OGvsYG et YGvsSD, on observe que les plants greffés présentent des DEGs majoritairement sur-exprimés par rapport à l'arbre adulte et majoritairement sous-exprimés par rapport aux plants semés.

J'ai ensuite associé chaque gène à un code BIN obtenu dans un fichier d'annotations (Perrin *et al.*, 2020) du pommier qui classe tous les gènes du pommier. Par exemple les gènes ayant le code BIN 1 sont associés à la photosynthèse.

J'observe dans un premier temps que le nombre de gènes associés à un code BIN (hors non annoté : code 35) est environ égal à la moitié des DEGs calculés pour chaque comparaison (Figure 4.2-8A), ce qui indique qu'une partie des DEGs n'est pas classifiée.

Pour la comparaison SDvsOG, 6 catégories représentent plus de 5% des DEGs, 7 catégories pour YGvsSD et 9 catégories pour OGvsYG. Je retrouve 4 catégories similaires dans chacune de ces trois comparaisons. Les catégories du métabolisme des lipides (13% de DEGs dans YGvsSD, 13,3% dans SDvsOG et 16,6% dans OGvsYG), celle du métabolisme des glucides (11,8%, 11% et 5% des DEGs respectivement), celles des acides aminés (7,6%; 8,9%; 8,4% des DEGs respectivement) et la catégorie correspondant au métabolisme des acides nucléiques (5,9%; 6,9%; 6,1% des DEGs respectivement).

Pour les DEGs communs entre deux comparaisons je retrouve trois catégories en communs (Figure 4.2-8B). La première correspond au métabolisme des lipides précédemment observé (24% des DEGs en communs entre YGvsSD et OGvsYG ; 10,5% entre YGvsSD et SDvsOG; 18,6% entre OGvsYG et SDvsOG). Je retrouve également la catégorie du métabolisme des glucides (10,6%; 11,4%; 6,8% des DEGs en communs, respectivement) et celle du métabolisme des acides nucléiques (5%; 5,6%; 8,1% des DEGs en communs, respectivement).

Enfin pour les DEGs en communs entre les trois comparaisons (Figure 4.2-7B) je trouve majoritairement des DEGs associés à la photosynthèse (20%) et à la respiration cellulaire (14%), ensuite je retrouve les quatre catégories citées précédemment : le métabolisme des glucides, des acides aminés, des lipides et des acides nucléiques représentant respectivement 7,6%, 7,2% et 6,3% et 5,7 % des DEGs en communs entre les trois comparaisons.

Nous avons donc des catégories de gènes différentiellement transcrits qui diffèrent entre les comparaisons et certaines qui sont communes. Ces catégories communes peuvent refléter des différences liées au mode de multiplication utilisé. Nous allons ensuite chercher s'il existe également un lien entre mode de multiplication et méthylation de l'ADN.

c) Régions différentiellement méthylées

Dans l'article présenté précédemment, nous avons analysé le nombre de DMRs calculé dans les comparaisons de l'arbre adulte versus les plants greffés (OGvsYG) et semés (SDvsOG) (Figure 4.2-9A). Ici je vais présenter les DMRs identifiés au sein de la comparaison entre les plants greffés et semés (YGvsSD), (Figure 4.2-9B) et observer s'il y a des différences avec ce que nous avons trouvé précédemment pour les comparaisons SDvsOG et OGvsYG.

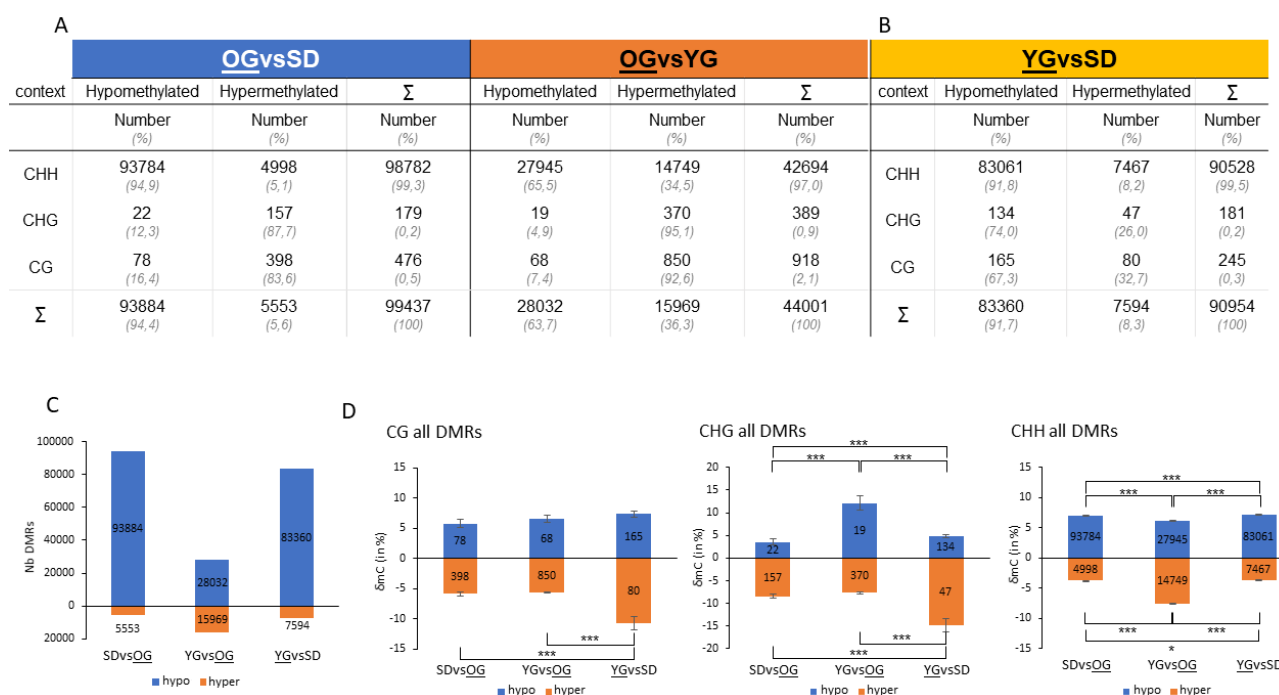


Figure 4.2-9 : Nombre, répartition et taux de méthylation des DMRs calculées dans chaque comparaison : DMA. Adapté de l'article Perrin et al. 2020. B. Table présentant la comparaison YGvsSD. C. nombre de DMRs hypométhylées (hypo ; bleu) et hyperméthylées (hyper ; orange) dans les trois comparaisons (Adapté de Perrin et al. 2020). D. Différence de méthylation (delta) entre les DMRs hyperméthylées ou hypométhylées pour chaque comparaison (Adapté de Perrin et al 2020). L'échantillon de référence pour le calcul des DMRs hyperméthylées ou hypométhylées est souligné pour chaque comparaison.

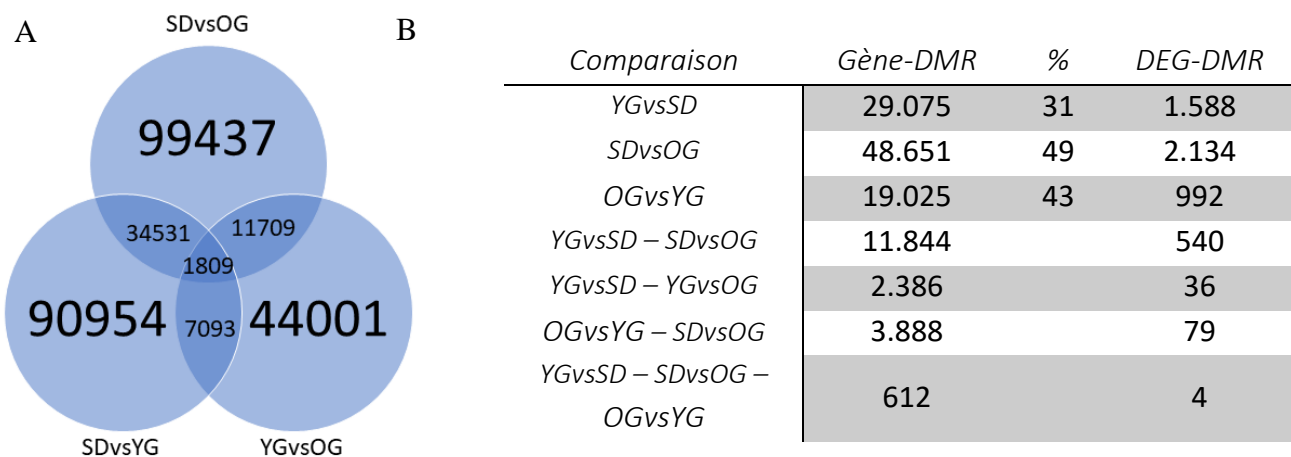


Figure 4.2-10: DMR entre les comparaisons et DEG DMR. A. A. DMRs communes tout contexte confondu. B. gène DMR et DEG DMR tout contexte confondu.

Nous avons identifié 90.954 DMRs pour la comparaison SDvsYG. La majorité de ces DMRs (99,5%) sont retrouvés en contexte CHH, 0,3% en contexte CG et 0,2% en contexte CHG. On retrouve un nombre de DMRs similaire à ce qui était observé dans la comparaison SDvsOG (99.437 DMRs) mais assez différent du nombre de DMRs dans la comparaison OGvsYG (44.001 DMRs).

On peut également noter que la répartition globale des DMRs hyper- et hypo-méthylées, est similaire entre les comparaisons SDvsOG et YGvsSD avec une grande proportion (plus de 90%) de DMRs hypométhylées dans les échantillons issus de semis par rapport à l'arbre adulte et au jeunes greffon, respectivement (Figure 4.2-9C) alors que dans la comparaison OGvsYG plus de 65% des DMRs sont hypométhylées dans l'échantillon YG par rapport à OG.

Je remarque également que les DMRs sont majoritairement hypométhylées dans les plants issus de semis (Figure 4.2-9A et B) dans les comparaisons SDvsOG et YGvsSD (94,4% et 91,7% de DMRs hypométhylées dans les plants semés, respectivement). Dans la comparaison OGvsYG, 36,3% des DMRs sont hyperméthylées dans YG par rapport à OG contre 91,7% de DMRs hyperméthylées dans YG par rapport à SD dans la comparaison YGvsSD.

Dans les comparaisons SDvsOG et OGvsYG j'observe que les DMRs en contextes CG et CHG sont hyperméthylées dans SD et YG par rapport à OG, avec 83,6% de DMRs hyperméthylées en contexte CG et 87,7% en contexte CHG pour la comparaison SDvsOG et 92,6% et 95,1% respectivement pour la comparaison OGvsYG.

En observant les comparaisons OGvsYG et YGvsSD, je m'aperçois que les DMRs en contextes CG et CHG sont plus méthylées dans les plants greffés, avec 92,6% et 95,1% respectivement pour la comparaison OGvsYG et 67,3% et 74% respectivement pour la comparaison YGvsSD.

Afin d'identifier si les DMRs calculées présentent une différence d'intensité en fonction du mode de transmission (sexuée ou asexuée) j'ai ensuite calculé les différences de méthylation moyennes dans chaque groupe (n.b. hyper- ou hypométhylé) pour chaque contexte de méthylation et pour chaque comparaison. Pour le contexte de méthylation CG je n'avais pas observé de différence de méthylation entre les comparaisons SDvsOG et OGvsYG (Figure 4.2-9D, Perrin et al). J'observe ici que le pourcentage de variations du taux de méthylation (ou delta de méthylation) est plus important

pour les DMRs hyperméthylées dans SD par rapport à YG en contexte CG (10,7%) que pour les DMRs hyperméthylées dans SD et YG par rapport à OG (5,9 et 5,6% respectivement). On notera également que l'effectif de DMRs hyperméthylées dans la comparaison SDvsYG en contexte CG (80 DMRs) est plus faible que dans SDvsOG (398 DMRs) et OGvsYG (850 DMRs).

Pour les contextes CHG et CHH on observe que les deltas de méthylation calculés, pour les DMRs hyperméthylées et hypométhylées, sont toujours significativement différents (p -value $< 1\%$) des deltas calculés pour les comparaisons SDvsOG et OGvsYG. On remarque néanmoins une différence significative, plus faible (p -value $< 5\%$) entre SDvsOG et YGvsSD pour les DMRs hyperméthylées en CHH.

Enfin j'ai comparé les DMRs calculées pour chaque comparaison afin d'identifier les DMRs communes entre comparaisons (Figure 4.2-10A). J'ai ainsi identifié 34.531 DMRs communes entre les comparaisons SDvsOG et YGvsSD, 11.709 DMRs entre SDvsOG et OGvsYG et 7.093 DMRs entre YGvsSD et OGvsYG. Enfin j'ai identifié 1.809 DMRs présentent dans chacune des trois comparaisons. On constate donc que l'échantillon SD semble être assez différent des arbres adultes et des greffons. De plus le nombre de DMRs commun entre SDvsOG et YGvsSD semble indiquer qu'il y a une régulation du niveau de méthylation pour certaines régions en fonction du type de multiplication. Ainsi la multiplication sexuée apporte beaucoup de variation alors que la propagation asexuée semble limiter la variation du méthylome.

d) Régions différentiellement méthylées en proximité des gènes

J'ai ensuite observé les DMRs proches des gènes dans chaque comparaison. C'est-à-dire les DMRs situées à moins de 2.000pb d'un gène et dans le corps des gènes. Dans un premier temps, je n'ai considéré que la présence d'un gène à proximité des DMRs (Gène-DMRs). Dans un second temps j'ai également pris en compte le niveau de transcription de ces gènes (DEG-DMRs) (Figure 4.2-10B). J'ai ainsi observé 48.651 DMRs proches d'un gène pour SDvsOG, 29.075 Gène-DMRs pour YGvsSD et 19.025 Gène-DMRs pour OGvsYG. Une partie de ces Gène-DMRs est commune entre les comparaisons, 612 Gène-DMRs sont retrouvés dans les trois comparaisons, 11.844 Gène-DMRs entre SDvsOG et YGvsSD, 3.888 Gène-DMRs entre SDvsOG et OGvsYG et 2.386 Gène-DMRs entre OGvsYG et YGvsSD. On constate que dans la comparaison SDvsOG les DMRs semblent être plutôt proche des gènes puisque 49% des DMRs sont proches des gènes contre 43% pour OGvsYG et 39% pour YGvsSD.

Pour l'observation des DEG-DMRs j'ai supprimé les doublons des gènes ayant plusieurs DMRs proches pour ne conserver qu'une seule occurrence de chaque gène. Ainsi j'ai calculé 2.134 DEG-DMRs pour SDvsOG, 1.588 DEG-DMRs pour YGvsSD et 992 DEG-DMRs pour OGvsYG.

Une partie des DEG-DMRs sont également retrouvée en commun entre les comparaisons, notamment 540 DEG-DMRs entre SDvsOG et YGvsSD, 79 DEG-DMRs entre SDvsOG et OGvsYG et 36 DEG-DMRS entre OGvsYG et YGvsSD.

Les DEG-DMRs retrouvées dans la comparaison YGvsSD représentent 5,15 des gènes annoté dans le génome de GDDH13, 6,55% pour la comparaison SDvsOG et 2,64% pour la comparaison OGvsYG.

Enfin j'ai identifié 4 DEG-DMRs (en contexte CHH) en communs entre tous les échantillons. Deux correspondent à des protéines de fonction inconnue (MD02G1204100 et MD17G1113100). Les

deux autres sont annotés AGAMOUS-like 20 pour la première (MD02G1197400) et comme appartenant à la super famille des Alpha/Beta hydrolase pour la seconde (MD12G1032200).

4.2.2. Comparaison de deux méthodes de calculs utilisées pour l'analyse des DMRs

A partir des séquençages bisulfites réalisés précédemment j'ai effectué différents calculs de DMRs en modifiant certains paramètres du pipeline développé par Nicolas Daccord (publié sur la plateforme GitHub (<https://github.com/EtienneBucher/DAMOCLES>)) précédemment utilisé dans Perrin *et al.*, (2020). Je voulais ainsi identifier les paramètres importants pour le calcul de DMRs en fonction de l'étude que l'on souhaite réaliser.

a) Comparaison globale des deux méthodes de calculs

J'ai effectué deux principaux calculs de DMRs pendant ma thèse. La Table 4.2-2 résume les caractéristiques de chacune de ces deux méthodes. La première différence majeure à noter est la version du génome utilisée pour réaliser l'alignement des séquences issues du séquençage bisulfite. En effet la première méthode de calculs a été appliquée sur la version V1.0 du génome de référence alors que pour la seconde méthode de calculs, j'ai utilisé la version V1.1 du génome de référence. Les différences entre ces deux versions du génome ont été décrites par Nicolas Daccord dans son manuscrit de thèse (Daccord, 2018).

Ces deux méthodes diffèrent également par les filtres utilisés. En effet, avec la première méthode les DMRs ont été filtrées directement en sortie du pipeline et seules les DMRs ayant une p-value inférieure à 5% ont été conservées. Néanmoins je n'ai pas accès à la p-value de ces DMRs dans le fichier de sortie pour cette méthode.

Pour la seconde méthode, nous avons ajouté différentes valeurs associées à la qualité du calcul des DMRs dans le fichier de sortie. J'ai notamment eu accès à la p-value pour chaque DMRs. Ce qui m'a permis d'appliquer un filtre plus strict à 1% pour la p-value. Nous avons également ajouté la moyenne des écarts-types (« standard deviation average », SDA) qui permet de rendre compte de la variation du taux de méthylation entre les répliques biologiques et entre les échantillons comparés. Nous avons précédemment fixé des seuils pour la SDA (Perrin *et al.* 2019) que j'ai choisi de réutiliser ici, notamment pour la comparaison YGvsSD présentée ci-dessus.

Malgré ces filtres plus stricts, que nous avons appliqués pour la deuxième méthode (Table 4.2-2), j'ai identifié un plus grand nombre de DMRs proche des gènes qu'avec la première méthode. En séparant les contextes je me suis aperçu que la grande majorité de ces DMRs ont été retrouvée en contexte CHH.

Pour les contextes CG et CHG j'ai retrouvé moins de DMRs par la deuxième méthode de calcul que par la première. Et j'ai également observé que dans les comparaisons OGvsYG et SDvsOG les DMRs en contextes CG et CHG sont majoritairement hypométhylées dans YG et SD par rapport à OG pour la première méthode de calcul. Alors que dans la seconde méthode de calcul, les DMRs en contextes CG et CHG sont majoritairement hyperméthylées dans YG et SD par rapport à OG. Pour la comparaison YGvsSD j'ai observé cependant que les DMRs en contextes CG et CHG sont majoritairement hypométhylées dans SD par rapport à YG entre les deux méthodes de calculs.

Enfin j'ai comparé le nombre de gène différentiellement transcrit et comportant une DMRs dans leur région promotrice ou terminatrice ou, dans le corps du gène (DEG-DMR). Et j'ai pu constater que l'on retrouve beaucoup plus de DEG-DMRs avec la seconde méthode de calcul des DMRs qu'avec la première, quel que soit la comparaison considérée.

Table 4.2-2 : tableau récapitulatif des deux méthodes de calculs de DMRs utilisées. L'échantillon de référence pour le calcul des DMRs est souligné dans le nom des comparaisons

Version du génome	1 ^{er} méthode		2 nd méthode			
	GDDH13 V1.0		GDDH13 V1.1			
Découpage des DMRs	Fenêtre fixe autour et dans les gènes		Fenêtre glissante de 200pb sur tout le génome (chevauchement de 100pb entre chaque fenêtre),			
Filtres utilisés	<i>p-value</i> (5% par défaut dans le pipeline)		<i>p-value</i> : 1% SDA – déviation standard de la moyenne de méthylation CHH : 0.05 (OG, SD, YG) CG – CHG : 0.07 (OG, YG), 0.11 (SD) (cf perrin et al 2020)			
	Contexte	Hypométhylé	Hyperméthylé	Hypométhylé	Hyperméthylé	
Nombre de DMRs proches des gènes (dans les 2000pb autour des gènes)	<u>Adulte</u> vs Greffée	CG	503	109	16	445
		CHG	417	232	10	155
		CHH	28	102	11497	6902
	<u>Adulte</u> vs Semée	CG	1179	339	48	208
		CHG	1279	415	14	89
		CHH	127	52	46471	1821
	<u>Greffée</u> vs Semée	CG	301	253	71	21
		CHG	226	368	97	20
		CHH	33	298	27180	1686
Nombre de DEG-DMR	<u>Adulte</u> vs Greffée	CHH	42 (contexte CG et CHG compris)		992	
	<u>Adulte</u> vs Semée	CHH	190 (contexte CG et CHG compris)		2134	
	<u>Greffée</u> vs Semée	CHH	75 (contexte CG et CHG compris)		1588	

b) Première méthode : Découpage en régions autour des gènes

Pour cette première méthode, nous avons utilisé la V1.0 du génome. Les DMRs ont été calculées sur des segments d'ADN dans la région promotrice, la région terminatrice est dans le corps des gènes (Figure 4.2-11, méthode n°1). Les segments utilisés ont été définis par rapport au point d'initiation de la transcription et en direction de l'amont du gène pour le promoteur et à partir du point de terminaison de la transcription et en direction de l'aval du gène pour le terminateur. Nous avons utilisé des segments de 100, 200, 500, 1000 et 2000pb pour le calcul des DMRs dont le détail est disponible en Annexe 2.

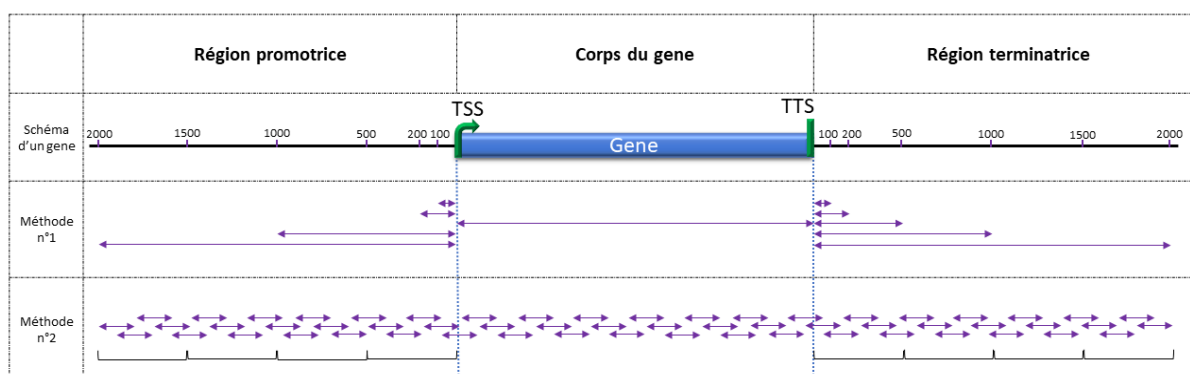


Figure 4.2-11 : schéma des différentes méthodes de calculs de DMRs à proximité d'un gène. Représentation d'un gène et de la position des DMRs calculées en fonction des différentes méthodes de calculs. Les flèches violettes représentent les DMRs, leur longueur représente les longueurs des DMRs calculées. TSS: site d'initiation de la transcription. TTS: site de terminaison de la transcription.

De façon assez intuitive, nous avons trouvé moins de DMRs pour les segments définis sur une courte distance par rapport aux plus grands segments définis (Table 4.2-3, voir aussi l'Annexe 2 pour les résultats plus détaillés).

D'autre part le nombre de DMRs calculé dans le corps du gène reste inchangé en fonction de la longueur du promoteur ou du terminateur considérée pour chaque calcul au sein d'une même comparaison. Nous observons seulement une variation de quelques DMRs (entre 1 et 8 DMRs de différence) dans la comparaison SDvsOG entre les calculs pour 100 et 200pb. Cette différence étant probablement liée à une erreur lors de l'analyse des données, cette variation sera sans conséquence pour la suite de notre analyse.

J'ai observé une tendance à une sous-représentation du contexte CHH (Table 4.2-3) dans les DMRs proches des gènes, quelle que soit la comparaison et la longueur de calculs des DMRs considérées. Par exemple pour la comparaison OGvsYG, pour les DMRs calculées sur 1000 pb dans la région promotrice, il y a 127 DMRs en contexte CG, 123 en contexte CHG et 28 en contexte CHH. J'observe également que le nombre de DMRs entre les contextes CG et CHG est relativement proche.

Étant donné le faible nombre de DMRs en contexte CHH proche des gènes et les similarités entre CG et CHG, j'ai décidé pour la suite de regrouper l'ensemble des DMRs CG, CHG et CHH pour simplifier l'analyse. A partir de cette liste de DMRs, j'ai observé la répartition du nombre de DMRs en fonction de leur localisation par rapport aux gènes et de la taille de ces DMRs (Figure 4.2-12A et B).

J'ai observé dans un premier temps une grande différence au niveau du nombre de DMRs hyperméthylé ou hypométhylé dans l'arbre adulte dans les comparaisons SDvsOG et OGvsYG. Par exemple, pour des DMRs d'une longueur de 2000 pb dans la région promotrice des gènes, il y a 920 DMRs hypométhylées dans SD par rapport à OG et 333 dans YG par rapport à OG pour 265 DMRs hyperméthylées dans SD par rapport à OG et 155 dans YG par rapport à OG. Pour la comparaison YGvsSD on observe qu'il y a plus de DMRs hyperméthylées dans SD(327) que de DMRs hypométhylées (217) par rapport à YG.

Table 4.2-3 : nombre de DMRs calculé avec la première méthode de calculs. Nombre de DMRs par comparaison en fonction du contexte et de la longueur de la DMR calculée. Voir tableau en Annexe 2 pour des résultats détaillés.

Contexte	Longueur	CG			CHG			CHH		
		SDvsOG	OGvsYG	YGvsSD	SDvsOG	OGvsYG	YGvsSD	SDvsOG	OGvsYG	YGvsSD
Region promotrice	2000	550	229	211	584	213	227	51	46	106
	1000	342	127	128	357	123	141	26	28	60
	500	178	62	73	190	75	87	12	14	39
	200	75	24	32	86	23	30	5	5	17
	100	48		12	45		15	4		6
Corps du gene	Body	184	63	83	190	80	93	50	30	57
	100	20		5	24		5	6		2
Region terminatrice	200	38	21	17	36	28	16	8	7	12
	500	106	33	31	118	46	37	11	7	23
	1000	210	90	71	246	114	69	22	15	44
	2000	392	160	130	460	178	137	39	27	84

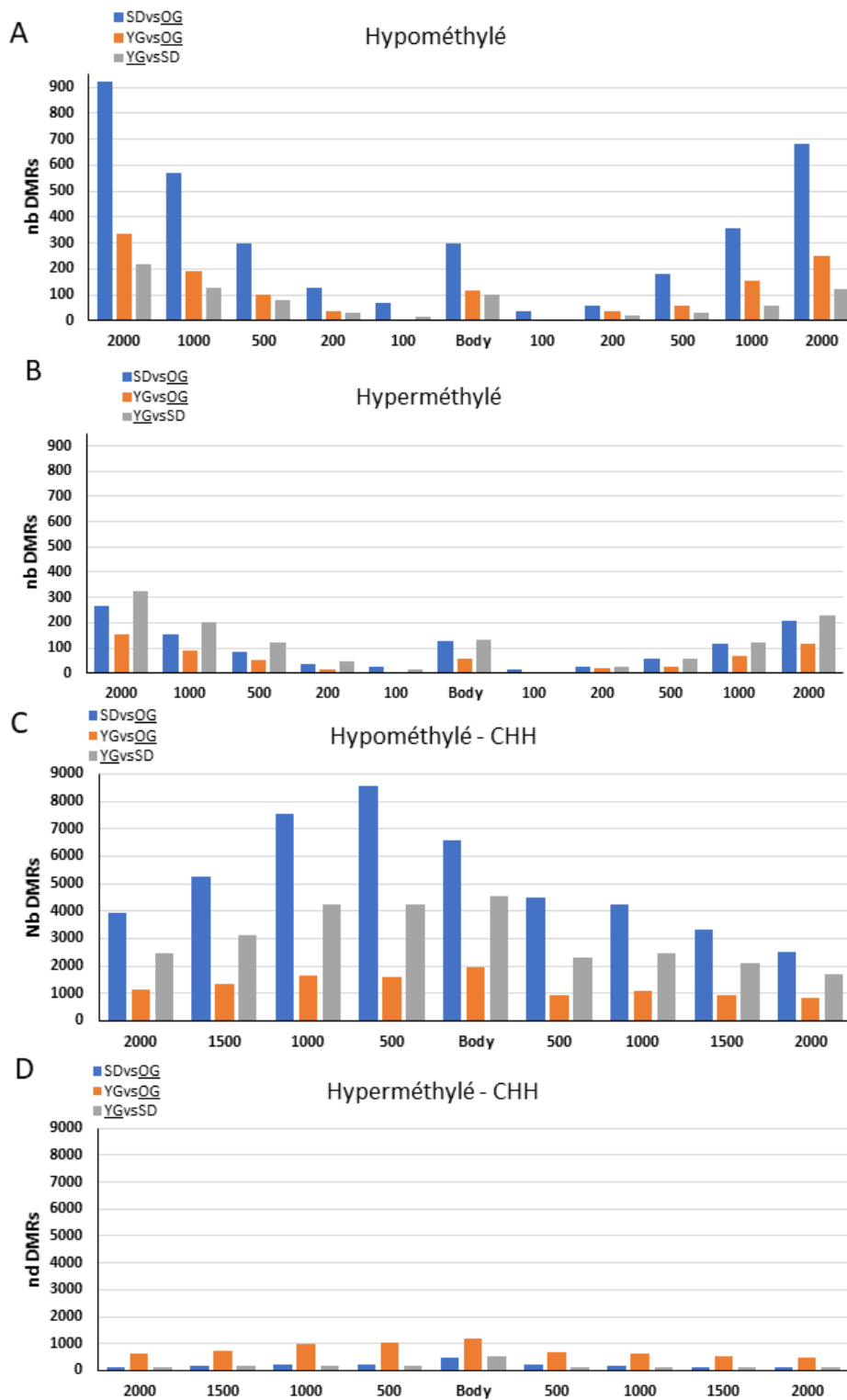


Figure 4.2-12 : répartition des DMRs par rapport aux gènes. A et B. nombre de DMRs calculé par la première méthode en fonction de la longueur de la DMRs dans les trois comparaisons considérées. DMRs hypométhylée en A et hyperméthylée en B. C et D. nombre de DMRs calculé par la seconde méthode en fonction de la distance de la DMRs par rapport aux gènes.

c) Seconde méthode : Sliding windows 200pb

Pour cette seconde méthode, nous avons utilisé la V1.1 du génome. Nous avons utilisé cette méthode dans l'article (Perrin *et al.*, (2020) pour l'analyse de la transmission de marque épigénétique chez le pommier.

Le nombre de DMRs et leur répartition en fonction des contextes et des comparaisons est disponible en figure Figure 4.2-9A qui est adapté de l'article Perrin *et al.*, (2020).

Pour comparer la répartition des DMRs dans les régions promotrices et terminatrices des gènes j'ai calculé le nombre de DMRs hyper- ou hypométhylé pour chaque comparaison en divisant ces régions en quatre parties égales (Figure 4.2-12C et D).

Tout comme dans la première méthode de calcul j'ai observé un plus grand nombre de DMRs dans la comparaison SDvsOG par rapport aux autres comparaisons. Si l'on considère les DMRs hypométhylés dans SD et YG par rapport à OG et dans SD par rapport à YG, entre 0 et 500pb dans la région promotrice du gène, je trouve 8565 DMRs pour la comparaison SDvsOG, 1608 pour la comparaison OGvsYG et 4225 pour la comparaison YGvsSD. J'ai également remarqué que contrairement aux précédents calculs, il y a moins de DMRs dans la comparaison OGvsYG que dans la comparaison YGvsSD.

Considérant maintenant la répartition des DMRs dans les régions promotrices et terminatrices des trois comparaisons (Figure 4.2-12 C et D). J'observe une différence de « densité » des DMRs. En effet il y a plus de DMRs proches du site d'initiation de la transcription et du site de terminaison de la transcription que dans les régions plus éloignées. Par exemple, pour les DMRs hypométhylés dans SD par rapport à OG, il y a 8565 DMRs entre 0 et 500pb dans la région promotrice des gènes alors qu'il y a moins de 4000 DMRs (3931 DMRs) entre 1500 et 2000pb dans la région promotrice. Il semble donc que la méthylation soit moins bien conservée à proximité de site d'initiation de la transcription des gènes. Ce phénomène est très marqué pour la multiplication sexuée (SDvsOG). On l'observe également dans la propagation asexuée (OGvsYG) mais dans une moindre mesure. On remarque également que ce phénomène semble absent de la comparaison YGvsSD.

Tout comme dans la première méthode, on retrouve également un nombre bien plus faible de DMRs hyperméthylés dans SD et YG par rapport à OG et dans SD par rapport à YG.

On remarque également que le nombre de DMRs hyperméthylé dans YG par rapport à OG est supérieur à celui des DMRs hyperméthylé dans les comparaisons SDvsOG et YGvsSD (Figure 4.2-12 D) contrairement à ce que nous avons observé pour la première méthode de calcul, quelle que soit la position des DMRs dans les régions promotrices et terminatrices ou dans le corps du gène.

4.3. Discussion

Dans cette première partie de ma thèse nous avons cherché à comprendre comment les marques de méthylations sont transmises lors d'une multiplication sexuée et lors d'une propagation asexuée. Nous avons pour cela utilisé un modèle d'étude unique, le pommier homozygote 'GDDH13' (Lespinasse *et al.*, 1999). Une partie de nos résultats a fait l'objet d'une publication dans laquelle nous avons notamment montré que le niveau de méthylation global reste changé à la suite de la multiplication de la plante (Voir figure 4A dans Perrin *et al.*, (2020)).

4.3.1. Effets de la transmission sexuée ou asexuée des marques épigénétiques

Nous avons pu le voir précédemment de nombreuses différences existent entre plants greffés et plants semés que ce soit au niveau phénotypique ou au niveau moléculaire.

a) Effets phénotypiques

Dans un premier temps j'ai recherché des différences phénotypiques entre les plants greffés et les plants issus de semis. J'ai observé des différences macroscopiques, comme le nombre d'étages foliaires (Figure 4.2-1), la forme des feuilles (Figure 4.2-3) la serration de la marge des feuilles (Figure 4.2-4) ou encore la densité de trichome sur la face inférieure des feuilles (Perrin *et al.*, 2020).

Concernant le nombre de nœuds, j'ai observé une différence au bout de cinq semaines de croissance. Les plants issus de semis présentaient alors un plus grand nombre de nœuds. Ceci permet de supposer que le développement des plants semés serait plus rapide que celui des plants greffés. Cet aspect pourrait être relié à un aspect de compétition existant dans la nature. En effet il a été montré que les plantes germant et se développant sous une canopée ont une croissance plus rapide que des plantes non soumises à l'ombre de la canopée (Valladares & Niinemets, 2008; Promis & Allen, 2017).

En outre la croissance d'un arbre diminue avec le temps une fois l'état adulte atteint (Bond, 2000). J'ai également pu montrer dans le chapitre précédent que les plants issus de greffes ne sont pas dans un état juvénile (Perrin *et al.*, 2020). Il est donc possible que les plants issus de semis présentent une réponse plus importante à l'ombre produite par les plantes voisines par rapport aux plants greffés et aient donc un développement plus rapide que ces derniers. Pour vérifier ceci il faudrait mettre en place des mesures de l'activité photosynthétique dans les différents échantillons.

J'ai ensuite comparé des feuilles d'âge différent prélevées sur les deux échantillons (Figure 4.2-3). J'ai pu identifier des différences au niveau de la forme des feuilles. J'ai trouvé que les feuilles issues de plants semés sont plus arrondies que les feuilles issues de plant greffés. Ce caractère a précédemment été identifié comme étant une différence observable entre des plantes juvéniles et des plantes adultes matures notamment chez *Olea europaea* L. (Basheer-Salimia, 2007).

J'ai également observé que le contour des feuilles issues de plants semés est plus linéaire avec des serrations régulières et bien marquées par rapport aux feuilles prélevées sur des plants greffés (Figure 4.2-4). Un tel développement de la marge des feuilles a été précédemment identifié comme reflétant un état juvénile de la plante échantillonnée (Karnatz, 1963; Hanke *et al.*, 2007).

De plus nous avons montré précédemment dans l'article Perrin *et al.*, (2020) que la densité de trichomes est plus faible chez les plants issus de semis que chez les plants greffés, ce qui est également un caractère lié à la juvénilité d'une plante (Chien & Sussex, 1996; Basheer-Salimia, 2007).

Les observations macroscopiques sont donc cohérentes entre elles. Elles confirment bien que les plants issus de semis sont dans un état juvénile. Cependant lorsque je confronte les observations macroscopiques des plants greffés à la littérature, je m'aperçois que les plants greffés ne peuvent pas être considérés comme des plantes dans un état juvénile.

Dans sa revue sur la juvénilité et la maturité des plantes, Basheer-Salimia, (2007) rappelle également l'existence de différences, au niveau microscopique, en termes de nombre de couches cellulaires ou d'organisation des parenchymes. Également une épaisseur plus importante des feuilles de plantes juvéniles par rapport aux feuilles de plantes matures a été constatée. Les coupes histologiques que j'ai réalisées (Figure 4.2-5) ne m'ont pas permis de vérifier ceci et les observations que j'ai effectuées, ne sont pas toutes concluantes. En effet, j'ai retrouvé une épaisseur du limbe (Figure 4.2-6B) des plants semés supérieure à celle des plants greffés. Cette différence a été évaluée par un test de Student et est significative à 5%. Mais la différence d'épaisseur est seulement de 4 μm , ce qui est inférieur à la taille d'une cellule. Cette mesure est donc peu concluante.

Concernant le diamètre de la nervure centrale (Figure 4.2-6B), celui-ci est supérieur dans les plants greffés par rapport aux plants semés (*p-value* pour le test de Student inférieur à 1%). Nous aurions pu nous attendre à l'inverse d'après la revue de Basheer-Salimia, (2007) et nos précédentes observations décrivant les plants greffés comme n'étant pas dans un état juvénile (Perrin *et al.*, 2020).

Enfin, sur l'ensemble des coupes histologiques, je n'ai pas observé de différences dans l'organisation des différents parenchymes et leur épaisseur contrairement à ce qui a été décrit et résumé précédemment dans la revue de Basheer-Salimia, (2007). Nous avons étudié ces différences sur des feuilles très jeunes. Les différences d'épaisseur du limbe apparaissent peut-être plus tard au cours de la croissance des feuilles.

Choix du phénotype pour la publication

Comme je l'ai décrit précédemment nous avons de nombreux caractères phénotypiques différents que nous aurions pu utiliser dans la publication. Cependant nous avons choisis de publier uniquement les résultats du phénotypage correspondant à la densité des trichomes. Nous avons fait ce choix car c'est un phénotype très simple à décrire et il a été facile de le convertir en données numériques pour réaliser des tests statistiques. Nous souhaitons plutôt nous concentrer sur les données qui nous paraissent les plus importantes pour cette publication, à savoir l'étude de la méthylation.

J'aurais aussi pu utiliser les différences de serration. Mais, si les serrations sont bien visibles et mesurables sur les plants issus de semis, il en va autrement pour les plants greffés, car les bordures de ces feuilles ont souvent tendance à s'enrouler et ne sont ainsi pas mesurables sur les photos. Je n'aurais pas pu réaliser des mesures du périmètre des feuilles de façon précise et reproductible.

Une évaluation de la forme des feuilles aurait également été possible, en effectuant un ratio largeur/hauteur. Mais la précision pour cette méthode aurait été là aussi très aléatoire.

b) Effet sur le niveau de transcription

Puisque nous avons identifiés des différences d'ordre phénotypique, nous avons décidé ensuite de réaliser une analyse du niveau de transcription des gènes entre chaque échantillons afin d'identifier des différentiellement transcrits pouvant expliquer les phénotypes observés.

Tout d'abord, j'ai observé une différence du nombre de transcrits différentiellement exprimés dans chaque comparaison. La comparaison YGvsSD présentait le plus grand nombre de DEGs.

Lorsque j'ai comparé les comparaisons deux à deux, j'ai retrouvé le plus grand nombre de DEGs, lorsque les plants semés étaient présents dans les deux comparaisons.

Enfin en analysant le nombre de transcrits sous- et sur-exprimés dans les comparaisons seules et à l'interface de chaque comparaison, j'ai observé un pattern de transcription : Il semble en effet que les arbres adultes présentent un niveau de transcription global plus faible que les plants semés et greffés et que les plants semés présentent un niveau de transcription global plus élevé que les deux autres échantillons. Nous avons également identifié des familles de gènes pouvant être reliées aux phénotypes observés, notamment les gènes associés à la photosynthèse (Perrin *et al.*, 2020).

La comparaison du niveau transcriptionnel des plants semés et greffés corrobore nos précédentes conclusions (Perrin *et al.*, 2020) basées uniquement sur les comparaisons SDvsOG et OGvsYG. Les plants greffés sont à l'interface entre des arbres adultes et des plants semés d'un point de vu transcriptionnel global.

c) Régions différentiellement méthylées

Afin de compléter notre analyse des différences entre plants greffés et plants semés, j'ai étudié les différences de méthylation existantes entre ces deux échantillons pour étudier si un liens existe avec les observations précédentes. Pour cela j'ai réalisé un calcul des DMRs en utilisant la seconde méthode de calcul (Figure 4.2-9) de la même façon que les DMRs calculées dans la publication.

J'ai constaté que le nombre de DMRs dans la comparaison YGvsSD est très similaire à celui observé dans la comparaison SDvsOG, de plus on retrouve un plus grand nombre de DMRs commun entre ces deux comparaisons qu'entre les autres comparaisons. Le dénominateur commun de ces deux comparaisons est les échantillons issus de semis. La multiplication sexuée engendre des variations locales de méthylation plus importantes que la propagation asexuée.

J'ai fait la même observation pour les Gène-DMRs et pour les DEG-DMRs, à chaque fois le nombre d'occurrences communes entre SDvsOG et YGvsSD est supérieur au nombre d'occurrences entre les deux autres comparaisons (SDvsOG et OGvsYG ou OGvsYG et YGvsSD).

Cependant pour les Gène-DMRs et les DEG-DMRs dans chaque comparaison, on peut observer cette fois que le nombre d'occurrences dans YGvsSD est intermédiaire entre SDvsOG et OGvsYG (Figure 4.2-10). Ceci indique que des variations du niveau de méthylation existent dans la propagation asexuée et que ces variations sont en moins grands nombres que dans la multiplication sexuée.

d) Est-ce que l'analyse de la comparaison YGvsSD corrobore les observations faites précédemment pour SDvsOG et OGvsYG ?

Au cours de ma thèse j'ai exploré différents aspects de la variabilité entre arbres adultes, plants greffés et plants semés. J'ai notamment exploré les marqueurs phénotypiques de la juvénilité chez le pommier tel que la serration de la marge des feuilles (Karnatz, 1963), la faible densité de trichome sur la face abaxiale des feuilles (Perrin et Al.) et l'absence d'initiation florale (Visser, 1964).

A travers la comparaison YGvsSD j'ai cherché à valider le modèle défini dans la publication précédemment présentée (figure 8 dans Perrin *et al.*, (2020)).

Ainsi les différents caractères phénotypiques observés, tels que la serration et la forme des feuilles chez les plantes juvéniles, n'ont pas été retrouvés chez les plants greffés (Figure 4.2-3, Figure 4.2-4). Mais tout comme les plants semés, les plants greffés ne sont pas capables d'initier la floraison avant la maturité complète, plaçant donc les plants greffés à l'interface entre juvénilité et maturité (Figure 4.3-1A).

L'analyse du niveau de transcription des gènes au sein de cette comparaison (Table 4.2-1 et Figure 4.2-7A) nous a permis l'observation d'une gradation du nombre de gènes différentiellement transcrits entre les différents échantillons, les plants issus de semis présentant le plus grand nombre de transcrits sur-exprimés, suivit des plants greffés (Figure 4.3-1B).

Enfin l'analyse de la méthylation de l'ADN pour cette comparaison, m'a permis d'affiner le modèle présenté dans la publication. En effet dans la publication nous avons observé que le niveau global de méthylation entre nos échantillons reste constant (Figure 4A dans perrin et al et Figure 4.3-1C.1.). Nous avons vu que les plants greffés se trouvent à l'intermédiaire entre l'arbre adulte globalement hyperméthylé au niveau des DMRs et les plants semés globalement hypométhylés au niveau des DMRs.

Ici j'ai pu définir que ce classement dépend de l'échelle d'analyse. En effet si l'on observe les différences de méthylation sur l'ensemble des DMRs (Figure 4.2-10A) on voit que les plants greffés semblent plus proches des arbres adultes (Figure 4.3-1C.2.) alors que nous aurions pu nous attendre à un nombre de DMRs intermédiaire pour YGvsSD entre le nombre de DMRs de SDvsOG et de OGvsYG, d'après le modèle défini dans la publication.

En se concentrant sur les régions DMRs proches des gènes uniquement, j'ai pu observer des différences plus importantes entre les arbres adultes et les plants greffés (Figure 4.2-10B). Ce qui permet ainsi de définir qu'à proximité des gènes les arbres adultes sont plus méthylés que les plants greffés qui sont eux-mêmes plus méthylés que les plants semés (Figure 4.3-1C.3.).

J'ai ainsi donc pu valider le modèle que nous avons publié précédemment (Perrin *et al.*, 2020). Enfin en utilisant la comparaison YGvsSD pour valider ce modèle, je montre qu'il est également possible d'utiliser les comparaisons SDvsOG, OGvsYG ou YGvsSD indifféremment pour ces analyses.

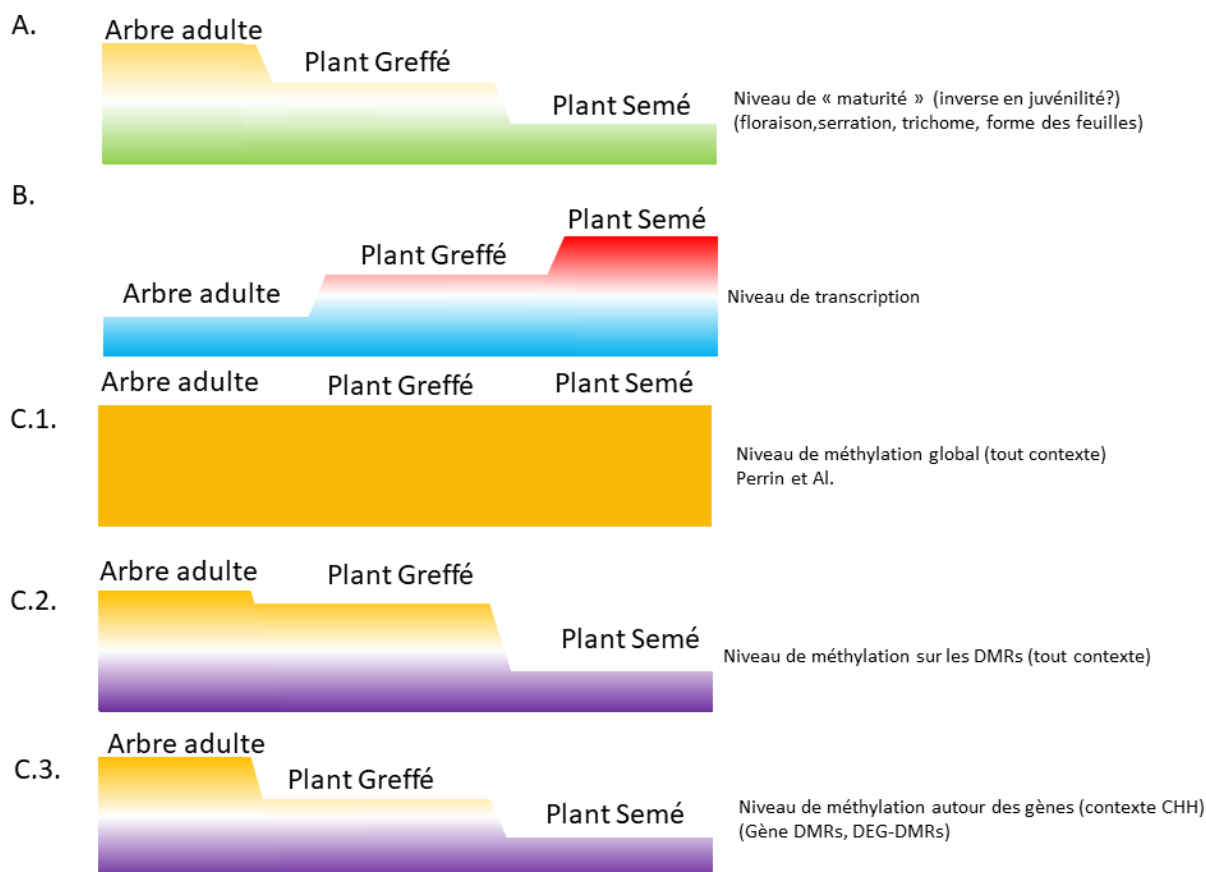


Figure 4.3-1 : Modèle général représentant les changements phénotypiques, transcriptomiques et épigénétiques ayant lieu lors d'une multiplication sexuée ou d'une propagation asexuée chez *Malus domestica*.

4.3.2. Discussion des méthodes de calculs des DMRs utilisées et proposition d'autres méthodes de calculs

A partir des données de séquençage bisulfite, j'ai réalisé des analyses comparatives du niveau de méthylation entre différents échantillons (un arbre adulte et des jeunes plants greffés et semés). Avec l'aide de Nicolas Daccord qui a développé un pipeline pour calculer les régions différentiellement méthylées entre différents échantillons (Daccord *et al.*, 2017) j'ai réalisé plusieurs calculs en modifiant certains paramètres. Le premier paramètre que nous avons modifié était la version du génome utilisée pour aligner les séquences. Dès que la V1.1 a été disponible il paraissait évident de l'utiliser puisque l'annotation et la couverture du génome sont meilleure que la V1.0. Cependant en changeant de version nous avons également modifié la méthode de calcul. Ce qui m'a amené à me questionner sur ces différents changements et ceux que l'on pourrait encore faire.

a) Était-il judicieux de faire un calcul sur le génome complet plutôt que sur des régions spécifiques ?

Nous avons donc réalisé deux calculs différents sur deux versions du génome différentes, alors que nous aurions pu utiliser la première méthode de calcul sur la V1.1 du génome. Nous avons fait ce choix pour plusieurs raisons.

Tout d'abord, du fait de l'expérience acquise dans l'équipe sur d'autres projets. Durant sa thèse, Nicolas Daccord a notamment réalisé une analyse comparative du méthylome de fruits et de feuilles. Il a ainsi pu définir que seule la méthylation sur la région située 100pb en amont du site d'initiation de la transcription permet d'observer une corrélation avec le niveau de transcription des gènes (voir figure 3.4 de la thèse Daccord 2018 en annexe 2). J'ai voulu réaliser la même analyse, mais le très faible nombre de DEG-DMRs que j'ai obtenu sur les segments de 100bp dans la région promotrice ne l'a pas permis.

Ensuite le calcul de DMRs pour les segments les plus longs ne rend pas compte des variations de méthylation pouvant exister sur le promoteur. Et les DMRs calculées sur 2000pb peuvent refléter en réalité une différence de méthylation sur quelques cytosines différenciellement méthylées seulement (seuil fixé à 5 cytosines différenciellement méthylées minimum). Alors que chaque segment de 2000pb est en moyenne constitué de 350 à 500 cytosines (méthylées ou non) en fonction de la proportion de GC (comprise entre 35 et 50% en moyenne chez les plantes, Wendel *et al.*, (2012)). Il y a donc là une imprécision importante dans le calcul de DMRs sur des segments si importants. Mais appliquer des filtres plus stricts auraient réduit le nombre de DMRs observé qui était déjà faible. De plus cette méthode tend à mésestimer le nombre de DMRs en contextes CHH. En effet nous aurions pu nous attendre à un nombre plus important de DMRs dans ce contexte comme observé dans d'autres expériences (Calarco *et al.*, 2012; Daccord *et al.*, 2017; Kawakatsu *et al.*, 2017; Xing *et al.*, 2019).

Enfin l'annotation des gènes entre les deux versions a été modifiée. Et je n'ai pas pu réaliser la correspondance de l'ensemble des gènes entre la V1.0 et la V1.1. Pour la V1.1 j'ai 51651 gènes et je n'ai retrouvé de correspondance dans la V1.0 que pour 24118 de ceux-ci, à partir de la correspondance avec les sondes MDP définies par (Velasco *et al.*, 2010).

Il était donc plus judicieux pour nous de réaliser un calcul de DMRs avec des fenêtres plus courtes (200pb) et chevauchantes. De plus l'annotation du génome (V1.1) était meilleure notamment pour les éléments transposables. Il était intéressant de réaliser un calcul complet de DMRs sur tout le génome et de filtrer ensuite sur les régions qui nous intéressent. Même si cela augmentait le temps de calcul de DMRs.

b) Pourquoi avoir choisi de faire les calculs avec la deuxième méthode plutôt que la première ?

Comme je l'ai déjà évoqué la V1.1 du génome présente une meilleure couverture et une meilleure annotation notamment des régions répétées et des centromères. Il était donc possible d'analyser les DMRs sur l'ensemble du génome de façon plus précise.

De plus l'utilisation du découpage en fenêtre coulissante de 200pb avec un chevauchement de 100pb offre une bonne résolution des différences de méthylation existantes entre deux échantillons. Avec une telle méthode on peut espérer expliquer une variation du niveau de transcription d'un gène par rapport à la structure d'une séquence couverte par une DMRs dans la région d'un gène ou à l'intérieur d'un gène. Notamment par l'identification de sites (ou d'un site) de fixation de facteur de transcription ou d'initiation de la transcription qui serait actif ou inactif en fonction du profil de méthylation de la séquence.

Enfin pour répondre à la question que je me suis posée dans la publication (et dans la thèse également), à savoir : comment sont transmises les marques épigénétiques entre une multiplication sexuée et une propagation asexuée, il était nécessaire de faire une analyse sans *a priori* pour limiter

les biais, hors par la première méthode on conserve un biais puisque nous n'avons travaillé que sur les gènes, ce qui ne représente que 23,3% du génome du pommier (Daccord *et al.*, 2017).

c) Quelles autres méthodes de calcul aurions-nous pu mettre en place et quelles seraient leurs avantages

Au fil de mes analyses du méthylome et de diverses discussions avec mes encadrants et d'autres chercheurs, j'ai réfléchi à d'autres façons de calculer les DMRs et de les appréhender. Je vais donc présenter quelques méthodes de calcul (Figure 4.3-2) qui pourraient apporter des informations complémentaires ou différentes.

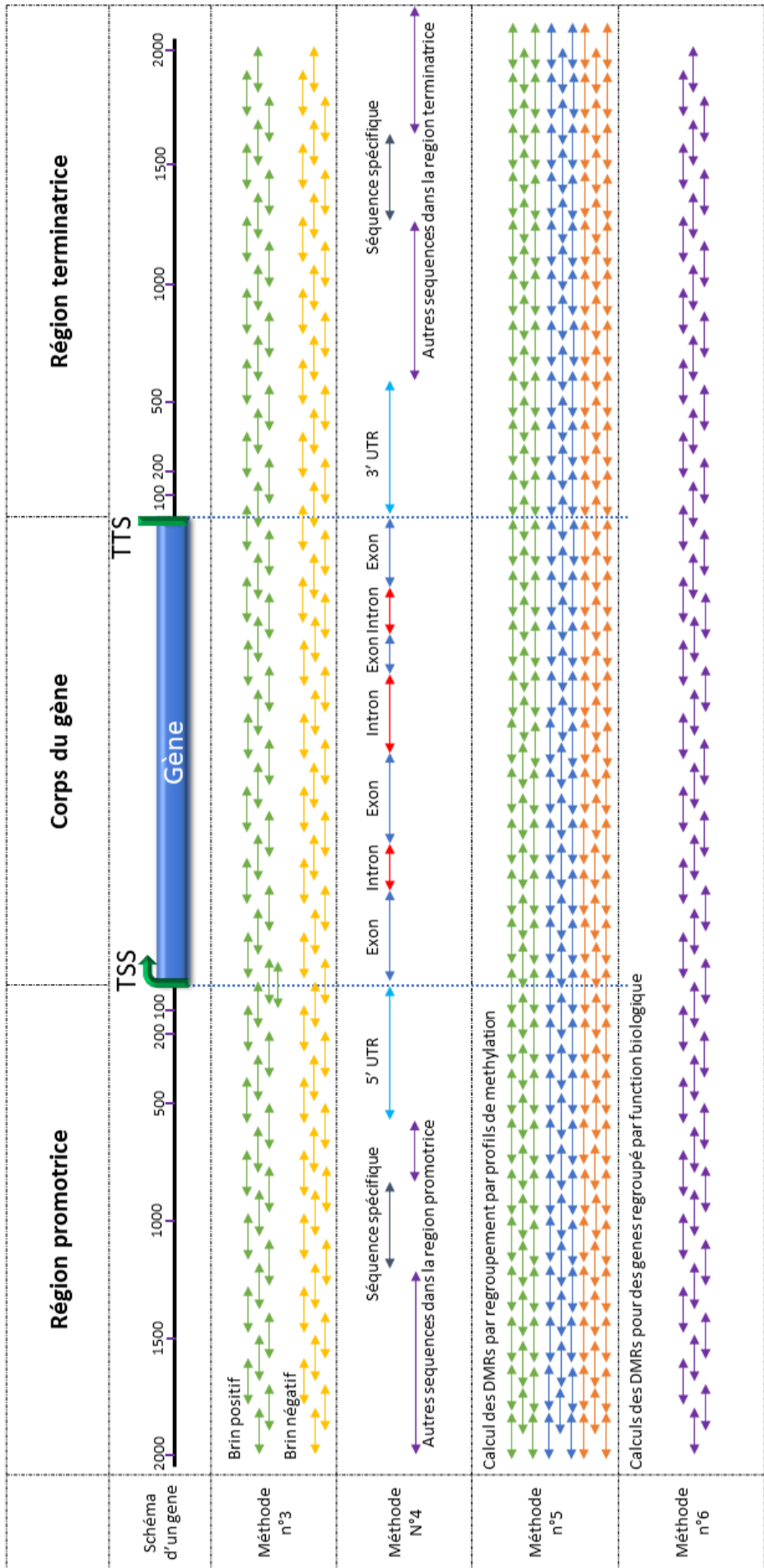


Figure 4.3-24-3-2: autres méthodes de calcul envisageables

Méthode 3 : Calcul des DMRs en fonction du brin

Cette troisième méthode de calcul est une variation de la seconde présentée précédemment (Figure 4.2-11 – méthode n°2). Ici les DMRs sont calculées en prenant en compte la localisation de la cytosine portant la méthylation sur le brin « positif » ou sur le brin « négatif » de l'ADN (Figure 4.3-2 – méthode n°3). J'ai imaginé cette méthode à la suite de visualisations de niveaux de méthylation sur le JBrowse. Je me suis aperçu qu'il apparaissait parfois des déficits de méthylation sur un brin, mais pas sur l'autre (Figure 4.3-3). Je me suis alors demandé l'effet que pourrait avoir ces déficits sur la transcription des gènes ou si certains déficits pouvaient être spécifiques à un échantillon. Et si finalement on pouvait identifier un nombre significativement différent de DMRs en fonction du brin considéré.

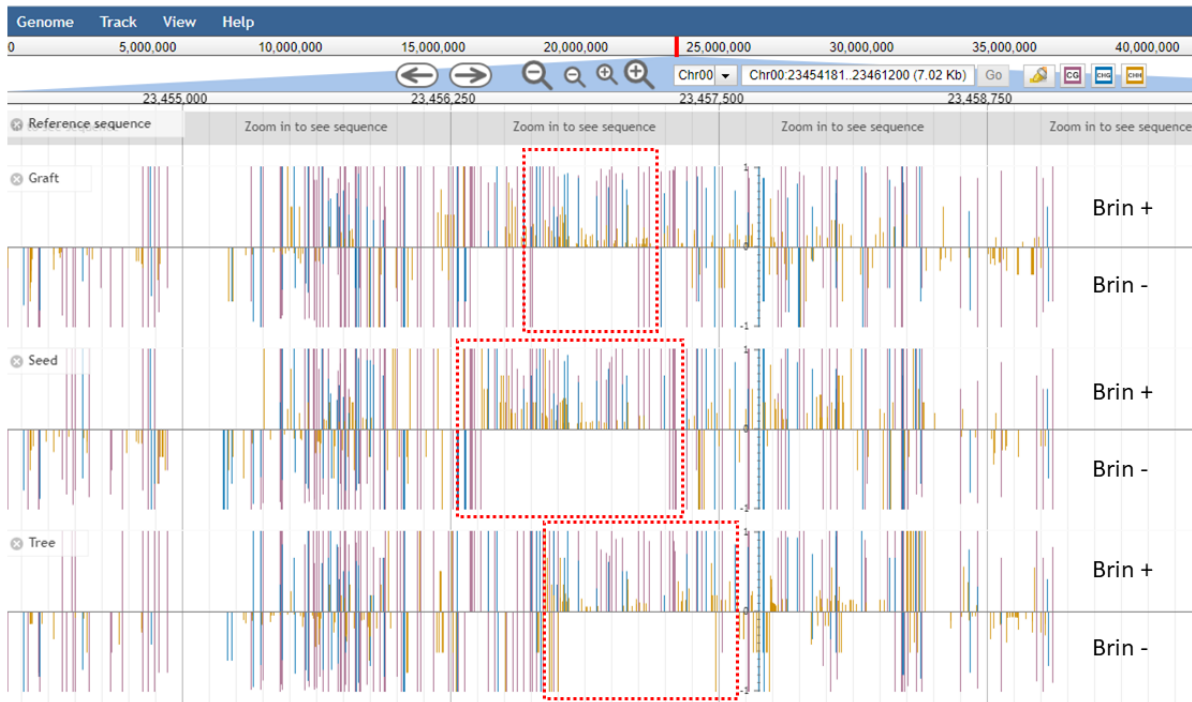


Figure 4.3-3 screenshot de JBrowse montrant un exemple de gap de méthylation. Brin + et - défini par convention sur le génome de référence. L'encart en pointillé rouge montre un gap de méthylation sur le brin - par rapport au brin + retrouvé quasiment à la même place entre les 3 types d'échantillon

Mais je n'ai pas pu appliquer cette méthode ici. En effet lors de l'assemblage du génome, les brins « positifs » ou « négatifs » sont définis de façon arbitraire pour chaque contig. Ainsi lorsque l'on passe d'un contig 1 à un contig 2, le brin « positif » du contig 1 peut devenir le brin « négatif » du contig 2.

Bien que ne limitant pas une analyse locale des DMRs, il est ainsi impossible de réaliser une analyse globale à l'échelle du génome comme je l'envisageais puisque l'on ne peut pas comparer les DMRs spécifiques d'un brin entre deux de contig.

Il faudra donc attendre un réassemblage du génome et l'obtention de 17 contigs correspondant aux 17 chromosomes du pommier pour effectuer une analyse globale. Dans l'état actuel de l'assemblage, seule une analyse avec *a priori* est envisageable. Par exemple pour analyser l'effet de la méthylation sur un cluster de gènes situé sur un même contig.

Méthode 4 : Calcul des DMRs en fonction et la structure des gènes

Pour cette quatrième méthode de calcul je propose de revenir à des fenêtres de calcul de DMRs fixes (Figure 4.3-2 – méthode n° 4). Les fenêtres seraient ici calculées en prenant en compte la structure des gènes (Figure 4.3-4) aussi bien pour les séquences intra-géniques (exons et intron(s)) que les séquences extra-géniques (5' et 3' UTR, site de fixation de facteur de transcription, îlots CG et autres séquences régulatrices en cis des gènes).

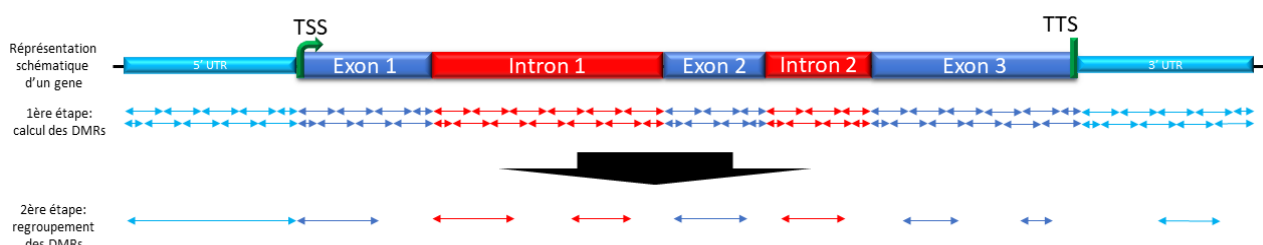


Figure 4.3-4: quatrième méthode de calcul: représentation schématique de la méthode

Pour éviter les problèmes liés aux calculs de DMRs sur de longues séquences et ne prenant en compte qu'un faible nombre de cytosine (problème soulevé pour la méthode de calcul n°1). Il est envisageable de réaliser un calcul de DMRs en scindant chaque séquence définie en sous séquence d'une longueur maximum de 200pb avec un chevauchement de 100pb. Une fois les DMRs ainsi calculées, toutes DMRs consécutives et présentant le même sens de régulation (*i.e.* hyperméthylation dans le même échantillon) seraient regroupées, de la même façon que présentées dans (Schalk *et al.*, 2016), pour n'en former qu'une au sein de la séquence initialement définie (*e.g.* : un exon). Il est à noter que cette possibilité de regrouper les DMRs consécutives est également implémentée dans le pipeline développé par Nicolas Daccord.

Cette méthode serait très intéressante pour réaliser des analyses ciblées sur un type de séquences dans l'ensemble du génome. Elle permettrait également une meilleure résolution pour l'étude du niveau de transcription des gènes en fonction de la position des DMRs.

Cette méthode présente cependant des limitations, notamment du fait de la très grande variabilité de la structure des gènes (*i.e.* : nombre exons et d'introns très variable d'un gène à l'autre). Il serait alors intéressant de séparer les gènes en catégories : petits gènes (de 1 à 4 exons), gènes de taille moyenne (de 5 à 9 exons), ou grands gènes (plus de 10 exons), par exemple. Pour faire cette classification il faut analyser la structure « moyenne » des gènes dans le génome (taille, nombre d'exons moyens, médiane du nombre d'exons). En 1999, Carels & Bernardi ont étudié la structure des gènes de 3 monocotylédones et de 6 dicotylédones. Ils ont notamment mis en évidence l'existence de deux classes de gènes chez les plantes, des gènes plutôt longs avec de nombreux introns, ayant la caractéristique d'être pauvres en GC et des gènes courts avec peu d'introns riches en GC. Nous pourrions nous appuyer sur cette classification pour réaliser des catégories de gènes.

Une seconde limitation est la nécessité d'avoir un génome bien annoté. Et également d'avoir une idée de la localisation des séquences régulatrices pour chaque gène.

Méthode 5 : prise en compte du profil de méthylation

Cette méthode est une application directe des analyses réalisées par Nicolas Daccord (2018) durant sa thèse (Figure 4.3-5A). Elle se base également sur les profils de méthylation observés pour les contextes CG, CHG ou CHH (Figure 4.3-5B) dans différentes espèces végétales telles que *Malus domestica* (Daccord *et al.*, 2017), *Arabidopsis thaliana* (Cokus *et al.*, 2008b; Hsieh *et al.*, 2009), ou *Picea abies* (Ausin *et al.*, 2016).

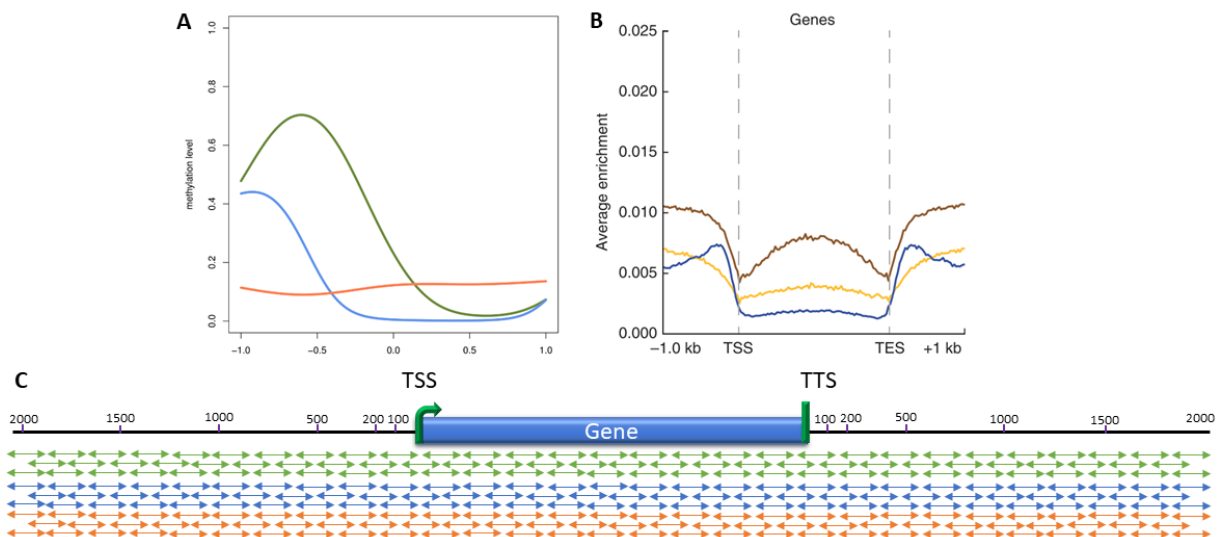


Figure 4.3-5: cinquième méthode de calcul: représentation schématique de la méthode. Adaptée de (Daccord *et al.*, 2017; Daccord, 2018). A. Regroupement de gènes en fonction d'un profil de méthylation commun (ici 3 groupes réalisés en contexte CHH), sur une distance de 100 pb avant le site d'initiation de la transcription. B. Niveau de méthylation moyenne pour les contextes CG, CHG, et CHH autour d'un gène. Adapté de Daccord, (2018). C. Schématisation de la méthode de calcul en fonction des données de la littérature. Les couleurs des doubles flèches représentent les clusters des gènes tels que défini en B.

L'idée est de regrouper les gènes méthylés en fonction des variations de la méthylation le long de leur séquence. Nicolas Daccord (2018) a réalisé ces regroupements en prenant 100pb avant le site d'initiation de la transcription. Il faudrait prendre des fenêtres plus grandes, jusqu'à 1000pb avant le site d'initiation de la transcription comme ce qui est fait pour l'analyse du niveau de méthylation autour des gènes en fonction des contextes (Figure 4.3-5B).

Pour cette méthode je propose ici un calcul des DMRs (en fonction du profil de méthylation) en utilisant la méthode des fenêtres glissantes de 200pb avec un chevauchement de 100pb (méthode 2, Figure 4.3-5C) mais limitée aux abords des gènes (+/-2000pb en amont ou en aval du gène). La méthode de calcul par fenêtres prédéfinies, en prenant en compte la structure du gène (méthode 4), serait également adaptée.

Les DMRs ainsi calculées seraient à relier aux gènes différentiellement transcrits, en séparant les DEGs en deux catégories de DEG : les gènes jamais méthylés et les gènes présentant une méthylation.

Méthode 6 : Analyse en fonction de la classification fonctionnel des gènes

Enfin pour cette dernière proposition de méthode pour calculer les DMRs (Figure 4.3-2), je propose de regrouper les gènes par catégories fonctionnelles (en utilisant l'annotation par code BIN utilisée dans le logiciel Mapman par exemple) (Lohse *et al.*, 2014).

Cette catégorisation pourrait permettre de mettre à jour des fonctions biologiques régulées par la méthylation de l'ADN. Je pense notamment au gène *FLC* associé à la vernalisation qui est réprimée épigénétiquement via la méthylation sur les histones (Berry & Dean, 2015). Ou à la répression des éléments transposables via la méthylation de l'ADN (Hollister & Gaut, 2009; He *et al.*, 2011).

Une fois la catégorisation faite un calcul de DMRs réalisé en utilisant des fenêtres glissantes de 200pb (méthode 2) ou en utilisant des fenêtres prédéfinies (méthode 4), devrait fournir de bons résultats.

Une telle étude demande d'avoir une caractérisation des gènes suffisante. Il est également nécessaire d'avoir un fichier d'annotations et de classifications des gènes de très bonne qualité (Cf : discussion sur la classification par code BIN, section 4.3.3 a).

d) Quelle méthode choisir ?

Comme nous l'avons vu pour les deux méthodes testées et pour les méthodes proposées, chaque méthode a ses avantages et ses inconvénients. Il convient donc dans un premier temps de bien définir quel type d'analyse nous souhaitons réaliser.

Pour une analyse à l'échelle du génome je recommande l'utilisation de la seconde méthode (fenêtres glissantes) qui est une analyse sans *a priori*.

Pour une analyse ciblée sur une séquence d'ADN défini à l'avance, comme un cluster de gènes, une analyse par fenêtre fixe définie selon la structure de la séquence d'intérêt me semble la plus appropriée.

La méthode idéale que j'aurais pu utiliser sur mes données regroupe les méthodes 3, 4, 5 et 6 (cette dernière pour une classification *a posteriori*). C'est-à-dire un calcul de DMRs avec des fenêtres glissantes de 200bp uniquement à proximité des gènes, en prenant en comptes la structure des gènes (exons et introns) et les profils de méthylation (région hyper-méthylée, pic de méthylation, méthylation moyenne constante). Avec ces résultats il devrait être plus facile de mettre en évidence des liens entre DMRs et niveau de transcription des gènes.

4.3.3. Critiques de différent points

Lors de la mise au point du plan expérimental, des expérimentations et des analyses nous avons fait des choix. Dans cette partie, je souhaitais revenir sur certains points afin de prendre du recul sur ce que nous avons fait : avons-nous fait les meilleurs choix ? Qu'aurions-nous pu faire différemment ?

a) L'utilisation de la classification par code BIN : une bonne idée ?

Lors de l'analyse des transcrits différentiellement exprimés, j'ai fait le choix d'utiliser un fichier de classification des gènes du pommier utilisables dans le logiciel MapMan®.

En analysant un peu ce fichier je me suis aperçu qu'il n'était pas parfait. En effet j'ai retrouvé des gènes ayant des fonctions telles que la photosynthèse (phytochromes ou portant la mention photosynthèse dans leur nom), dans la catégorie 35. C'est-à-dire la catégorie des gènes non classés. Lorsque j'ai fait des classifications de gènes (*i.e.* : Figure 4.2-7B et Figure 4.2-8) je n'ai pas pris en compte cette catégorie 35, qui représentait la plupart du temps plus de la moitié des DEGs. Je suis donc face à un biais potentiellement important ici. Mais je peux supposer que pour toutes les catégories, une partie des gènes pouvant y être associés se retrouve dans la catégorie 35. Et ceci de façon randomisée, réduisant ainsi le biais.

Il existe d'autres logiciels permettant de réaliser des catégorisations de gènes et également de l'ontologie de gènes, tel que agriGO® ou Panther®. Pour ces logiciels il faut en général passer par la correspondance des gènes avec *A.thaliana*. Ce qui, dans le cas de génome dupliqué comme le pommier, provoque un biais. Certains logiciels d'ontologie des gènes commencent à intégrer des fichiers d'annotations d'autres espèces végétales. C'est notamment le cas de agriGO®, qui intègre une annotation du pommier pour faire des analyses d'enrichissement de gènes. Cependant les identifiants des gènes sont des MDP correspondant au génome du pommier publié en 2010 (Velasco *et al.*, 2010). Et comme j'ai pu m'en apercevoir pendant ma thèse, la correspondance entre les annotations du pommier, n'est pas parfaite (cf partie 4.3.2. a)).

La classification des gènes n'est donc pas une chose facile à réaliser. On se heurte à de nombreux biais réduisant la puissance de notre analyse. Le fichier de classification que j'ai utilisé, même s'il n'est pas parfait, est néanmoins la meilleure solution que j'ai pu trouver, puisque cette classification a été réalisée à partir de la V1.1 du génome de référence tout comme le micro-array utilisé pour mon analyse de gènes différentiellement transcrits.

b) Discussion du plan expérimental

Nous avons mis en place un plan expérimental pour la production des échantillons et leurs prélèvements. Mais ce plan était-il le meilleur ? Aurions-nous pu réaliser des échantillons supplémentaires ou bien faire les prélèvements différemment ?

Pertinence du prélèvement la première année

Cette question de la pertinence du prélèvement pendant la première année de croissance se pose particulièrement pour les plants issus de greffes. En effet nous avons prélevé des baguettes portant des bourgeons sur un arbre adulte en verger. Ces bourgeons provenant d'un arbre adulte, peuvent être déjà induit ou quiescent. Ils peuvent donc, en principe, développer des organes floraux. J'ai pu observer ce phénomène, quelques semaines après le greffage en serre (Figure 4.3-6A), pour une partie des plants greffés. On peut observer cependant que les organes floraux apparaissent non pas sur du bois nouvellement formé, mais bien à partir des bourgeons prélevés sur l'arbre adulte.



Figure 4.3-6: Photos de pommier en serre et sous loupe binoculaire. A. Photographie d'un plant greffé. B et C: observation de feuille de pommier prélevé sur des plants semés (coin supérieur droit) et greffés (coin inférieur gauche) pour des plants obtenus en 2017. B. photo réalisée en 2017 sur des feuilles de l'année. C. Photo réalisée en 2018 sur des feuilles de l'année prélevées sur des plants de 1 an.

Il est donc raisonnable de se demander si le bois formé à partir de ces bourgeons a un épigénome identique à celui de l'arbre adulte ou si l'action du greffage et de la communication entre porte greffe et scion (Gregory *et al.*, 2013; Zhao & Song, 2014) a déjà eu pour effet de modifier l'épigénome, et notamment de diminuer la méthylation localement, comme nous l'avons montré précédemment (Perrin *et al.*, 2020).

Afin de limiter ce biais il aurait été préférable de laisser passer une année de croissance, de rabattre les arbres et de faire un prélèvement la seconde année afin d'avoir un épigénome de greffon « stabilisé ».

Ceci n'aurait pas eu d'incidence sur le phénotypage, comme le montrent les observations préliminaires que j'ai pu réaliser (Figure 4.3-6B et C). On peut observer une augmentation de la densité de trichomes sur les feuilles issues de semis de 1 an, mais cette densité reste toujours très inférieure à celle observée chez les plants greffés.

Pertinence du modèle double haploïde

Pour notre étude nous avons utilisé un matériel biologique très particulier puisqu'il s'agit d'un génotype de pommier homozygote (Golden delicious double haploïde, Lespinasse et al 1999). Mais ce choix était-il judicieux ?

Le double haploïde est un modèle d'étude simple

Comme présenté dans l'introduction, le pommier (*M. domestica*) est naturellement hétérozygote du fait de l'existence d'une auto-incompatibilité pollinique (Zhang *et al.*, 1988; Cornille *et al.*, 2014). L'étude de l'épigénome d'une plante hétérozygote est possible, mais complexe. Par exemple, il est difficile d'être certain qu'une variation phénotypique soit liée à une modification épigénétique et non pas à l'expression différentielle des allèles d'un même gène entre deux individus. L'étude de l'épigénome d'une plante homozygote est donc plus facile à réaliser, puisque la complexité du génome est réduite.

Mais finalement ce modèle homozygote ne correspond pas aux pommiers hétérozygotes retrouvés dans la nature et utilisés en verger. Il existe donc peut-être un biais dans nos résultats et notamment dans la façon dont les marques épigénétiques se transmettent via la multiplication sexuée et la propagation asexuée.

Conséquence de l'autofécondation

Là aussi nous sommes face à une question primordiale : pourquoi et comment avons-nous réussi à réaliser une autofécondation chez une espèce qui est normalement auto-incompatible. Daccord, (2018) explique qu'il est possible que, pendant les étapes d'obtention des doubles haploïde de Golden Delicious, des portions limitées du génome aient été perdues à la suite de réarrangements chromosomique. Un alignement de marqueurs SNP cartographiés sur des pommiers hétérozygote montre que la majorité des marqueurs (95,6%) sont retrouvés sur le génome de 'GDDH13' (Daccord *et al.*, 2017) étayant l'hypothèse de la perte de portion génique de façon limitée. Cette thèse est également étayée par des découvertes sur le genre *Arabidopsis*. Il a été montré que la mutation d'un gène impliqué dans la reconnaissance pollinique suffit à lever la barrière de l'incompatibilité pollinique (De La Chaux *et al.*, 2012).

S'il s'est passé la même chose pour 'GDDH13', cela pose question sur notre modèle. Puisque l'on touche au mode de reproduction, on peut se demander quelles sont les conséquences pour la transmission des marques épigénétiques via la multiplication sexuée ? Est-ce que cette transmission est similaire chez une espèce auto-incompatible et chez une espèce auto-compatible ? Il est donc possible que le modèle de transmission des marques épigénétiques entre multiplication sexuée et asexuée, que nous avons définie pour 'GDDH13', soit un peu différent de ce qui serait observable chez un *Malus domestica* hétérozygote.

5. Conclusion

Dans cette première partie nous avons analysé la différence entre une multiplication sexuée et une propagation asexuée chez le pommier. Nous avons ainsi pu mettre en évidence des différences aux niveaux phénotypiques que nous avons pu définir pour partie comme des marqueurs d'un état juvénile des plants issus de semis. Les jeunes plants obtenus par greffage présentent quant à eux un phénotype pseudo-mature.

Nous avons également identifié des différences transcriptionnelles existant entre les différents plants. Nous avons ainsi montré que les plants issus de semis présentent un niveau de transcription plus élevé que les arbres adultes. Là encore les jeunes plants issus de greffage présentent un niveau de transcription à l'interface entre les plants semés et l'arbre adulte.

Enfin nous avons la méthylation de l'ADN. Nous avons montré que d'un point de vue global il n'existe pas de différence de transmission de la méthylation entre multiplication sexuée et propagation asexuée chez le pommier. Nous avons cependant mis en évidence des différences de niveaux de méthylation locale (DMRs) d'autant plus marquées à proximité des gènes. Nous avons ainsi constaté que la méthylation est plus faible chez les plants issus de semis au niveau de ces DMRs par rapport à l'arbre adulte et aux plants issus de greffage. Cette méthylation plus faible à proximité des gènes semble être corrélée avec un niveau de transcription plus élevé chez les plants greffés, comme nous avons pu le montrer notamment pour les gènes associés à la photosynthèse. Les gènes différenciellement transcrits présentant également une DMRs représentent une très faible portion des gènes du génome du pommier.

En parallèle nous avons également proposé différentes méthodes afin de calculer les régions différenciellement méthylées entre différents échantillons. Nous avons ainsi défini que pour une analyse globale telle que celle que nous avons réalisée, une analyse sans *a priori*, en utilisant des fenêtres de 200 pb, semble être la meilleure méthode d'analyse.

Chapitre II – Edition du génome et de l'épigénome du pommier

Dans cette partie je présente le travail que j'ai réalisé dans le cadre du projet Genius (Genome ENgineering Improvement for Useful plants of a Sustainable agriculture) porté par différents laboratoires publics et des partenaires privés. L'un des objectifs majeurs de ce projet est d'expérimenter et d'acquérir de nouvelles techniques de génie génétique sur différentes espèces cultivées (tomate, pomme de terre, pommier, rosier, etc...), notamment pour l'édition du génome via la technique CRISPR-Cas9. Cette acquisition de connaissances est un enjeu majeur pour la recherche et la sélection variétale française. Dans ce cadre j'ai été impliqué dans le sous-projet BEEP (CRISPR-Cas9 Based Epigenome Editing in Plants).

1. Introduction

Le projet BEEP s'articule autour de deux objectifs sur le pommier dont le premier est de produire dans les fonds génétiques 'Gala' et GDDH13, des plantes transgéniques éteintes pour le gène *MET1*, de les phénotyper et de les génotyper. Le second objectif est de tester la possibilité de réaliser l'édition de l'épigénome du pommier et d'analyser ses conséquences sur la transcription du gène de floraison *MdTFL1*, utilisé ici comme gène rapporteur.

La mutagenèse dirigée du pommier par la technique CRISPR-Cas9 étant déjà maîtrisée au sein de mon laboratoire d'accueil, sur la variété 'Gala', j'ai pu me former à l'utilisation de cette technique. J'ai notamment utilisé ces connaissances pour réaliser l'extinction du gène SCOOP12 d'*A. thaliana* dans le cadre de la thèse de Kay Gully et publié dans l'article Gully et al. (2019). Par la suite, dans le cadre du projet BEEP, j'ai cherché à transférer cette technique de la variété 'Gala' vers le génotype GDDH13 afin de pouvoir réaliser des analyses de l'épigénome sur le génome de référence.

1.1. La technique CRISPR-Cas9

Cette technique est basée sur une découverte faite au sein du génome des bactéries (Barrangou *et al.*, 2007). Les CRISPRs (Clustered regularly interspaced short palindromic repeats) sont des petites séquences palindromiques répétées organisées en clusters. Ce système permet la réparation de l'ADN et une régulation génétique (Makarova, 2002; Guy *et al.*, 2004).

Au sein du génome des bactéries, les séquences CRISPRs sont retrouvées à proximité de gènes nommés CRISPR associated genes (Cas). Il a été proposé par (Makarova *et al.*, 2006) Makarova et al (2006) que l'association des séquences CRISPR et des gènes Cas est une adaptation du système de défense des bactéries. Les complexes CRISPR-Cas utilisent des ARN antisens qui sont produit à partir de l'insertion d'ancienne séquences ADN exogènes inséré dans le génome (Barrangou *et al.*, 2007; Orlova, 2009, Griffiths *et al.*, 2000).

En 2014, Doudna et Charpentier ont compilé les connaissances sur le sujet pour proposer une nouvelle méthode de génie génétique simple et peu couteuse permettant l'édition précise des génomes. On peut également noter le travail de synthèse de Cavanagh P. et Garrity A. publié sur le site : <https://sites.tufts.edu/crispr/> qui présente les origines et la domestication de ce système. Cet outil utilise des ARNs guides, conçus pour associer deux caractéristiques cruciales : une séquence ARN de 20 nucléotides en 5', qui détermine la séquence cible dans le génome par complémentarité de bases, et une structure ARN en 3' assurant la liaison à la Cas9. L'association ARN guide/Cas9 permettant à la Cas9 de réaliser une cassure double-brin site-spécifique au sein du génome.

1.1.1. Composition du système CRISPR-Cas9 et modalités de réparation des cassures double brin de l'ADN

Le système CRISPR-Cas9 est un complexe se composant des éléments suivants :

- Une enzyme associé au système CRISPR (Cas9). C'est une enzyme capable de couper l'ADN double brin grâce à des sites actifs appelés nucléases. Sur la séquence protéique de la Cas9, isolée de la bactérie *Streptococcus pyogenes*, deux sites nucléases ont été prédits par cristallographie (Nishimasu *et al.*, 2014), les domaines HNH et RuvC (**Error! Reference source not found.**A).

- un crRNA (ou CRISPR RNA) qui est une séquence ARN de 20 nucléotides complémentaire de la séquence d'ADN ciblée (Nishimasu *et al.*, 2014), (**Error! Reference source not found.B**).

- et enfin un tracrRNA (trans-activating crRNA, **Error! Reference source not found.C**, Deltcheva *et al.* 2011). Ce dernier est relié au crRNA et permet de donner la conformation nécessaire au couple crRNA+tracrRNA pour qu'il soit reconnu par la Cas9 (**Error! Reference source not found.D**). Par la suite l'association crRNA et tracrRNA sera appelée ARN guide (gRNA).

Le complexe ainsi formé peut aller cibler la séquence complémentaire du crRNA dans l'ADN (**Error! Reference source not found.E**). Une fois la séquence reconnue, les sites nucléases vont réaliser une coupure double brin de l'ADN, si la présence d'un PAM (Protospacer Adjacent Motif) adjacent à la séquence complémentaire au crRNA est bien reconnue. Le PAM est une petite séquence spécifique à chaque Cas, nécessaire au déclenchement de l'activité nucléase de cette dernière. Pour la Cas9, le PAM est de type « NGG » (où N peut être A, T, G ou C). Cette coupure double brin (coupure franche) sera ensuite détectée par les mécanismes de réparation de l'ADN de la cellule. Deux mécanismes coexistent : la réparation par recombinaison homologue (HDR, homology-directed repair) et la réparation non homologue (NHEJ, non-homologous end-joining) (voir Manova and Gruszka 2015 pour une revue complète). La réparation par recombinaison homologue est possible lorsque l'on apporte une séquence aux bordures homologues à celle ciblée par le système CRISPR-Cas9, qui sert de matrice pour la recombinaison (**Error! Reference source not found.**, Rong and Golic 2000).

Dans le cas de la réparation par NHEJ, aucun modèle n'est apporté pour réparer la cassure (**Error! Reference source not found.**). La machinerie de réparation cellulaire va alors rejoindre les deux extrémités de la cassure mais en créant potentiellement des erreurs par rapport à la séquence initiale (délétion, insertions, substitutions) et ainsi de la mutagenèse ciblée (Manova & Gruszka, 2015). Chez les plantes la réparation par NHEJ est beaucoup plus fréquente que la réparation par recombinaison homologue, ce qui participerait à l'augmentation de la taille des génomes ainsi qu'à leur complexification (Puchta & Fauser, 2014).

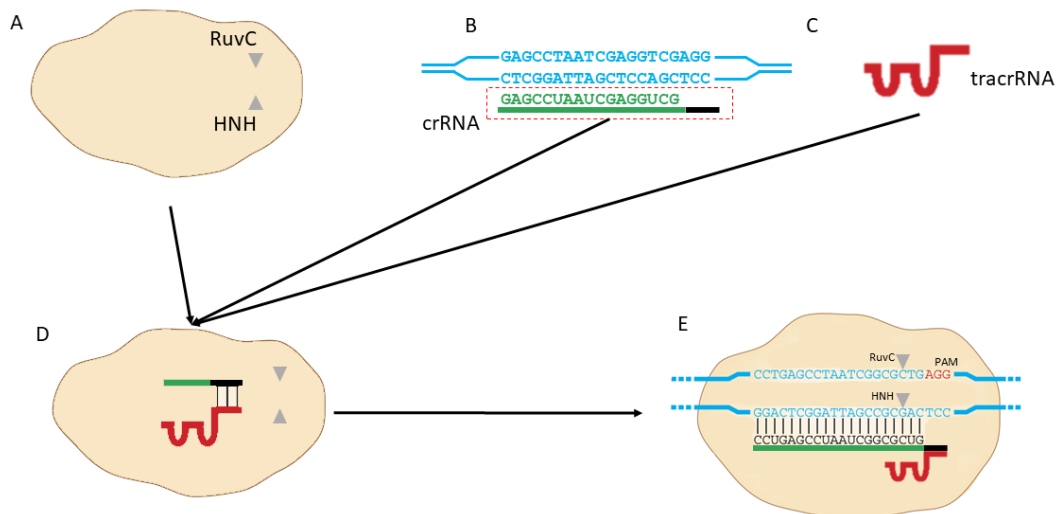


Figure 1.1-1 : Système CRISPR-Cas9. A. Représentation de la Cas9 contenant les deux sites nucléases. B. Représentation d'un crRNA complémentaire de la séquence d'ADN ciblée. C. tracrRNA, permettant le repliement de la structure crRNA+tracrRNA. Le tracrRNA permet la reconnaissance du complexe crRNA + tracrRNA et sa prise en charge par la Cas9 pour former le complexe final en D. E. le complexe Cas9 et sa cible. Adapté de Sander JD, Keith Joung J (2014) review CripSPr-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32:3-4. <https://doi.org/10.1038/nbt.2842>

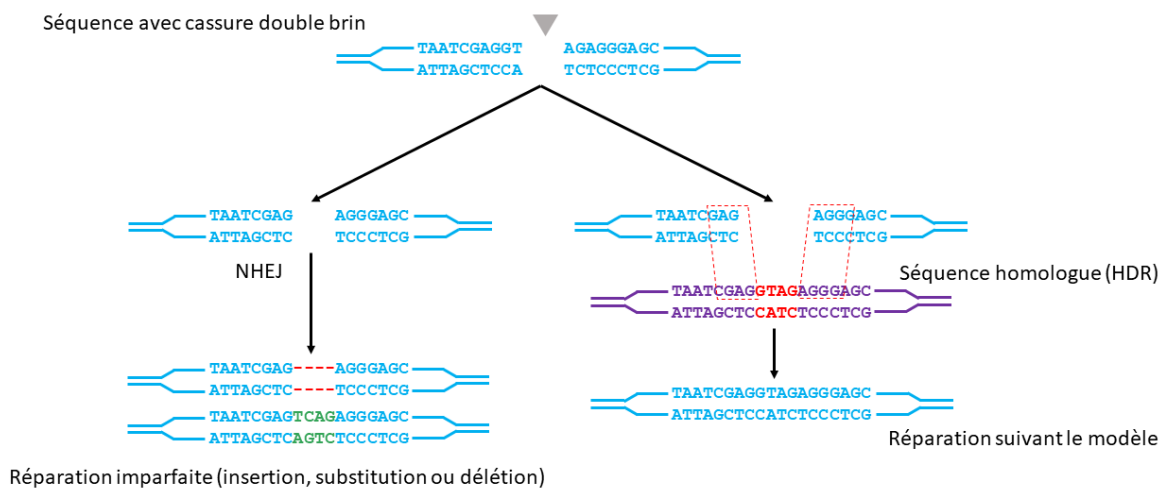


Figure 1.1-2: Schéma de réparation d'une cassure double brin via le NHEJ (à gauche) ou la HDR (à droite). D'après Manova V, Gruszka D (2015) DNA damage and repair in plants – From models to crops. Front Plant Sci 6:1-26. <https://doi.org/10.3389/fpls.2015.00885>

1.1.2. Evolution des applications de l'édition du génome

La mutagenèse dirigée grâce au système CRISPR-Cas9 a pu se développer grâce notamment à la découverte de nouvelles Cas dont les sites PAM, différents du NGG de la Cas9 de *Streptococcus pyogenes* (Shah *et al.*, 2013), permettent de cibler plus de séquences au sein des génomes.

Outre son utilisation pour de la mutagenèse dirigée, plusieurs initiatives ont émergé chez les plantes, ce qui augmente le champ d'action de cette technique, en particulier pour pallier au faible rendement de l'édition proprement dite des gènes, liés aux faibles taux de recombinaison homologue. C'est le cas du « base editing » (Bharat *et al.*, 2019) et du « prime editing » (Anzalone *et al.*, 2019).

Le « base editing » consiste à remplacer spécifiquement une base nucléotidique par un autre (Bharat *et al.*, 2019), de manière à modifier le gène cible pour lui conférer une nouvelle fonctionnalité. Cette technique est basée sur le même principe de reconnaissance de l'ADN ciblé par le système CRISPR-Cas9, mais il n'est plus question de faire de coupure double brin dans l'ADN. La Cas9 native est donc remplacée par une Cas9 totalement (« deadCas9 ») ou partiellement inactivée (« nickase » Cas9, nCas9) au niveau de ses sites nucléases. Cette Cas9 totalement ou partiellement inactivée est ensuite fusionnée à une désaminase permettant le remplacement de nucléotide. Par exemple une cytidine désaminase (Eid *et al.*, 2018) qui permet la transition d'un nucléotide C en T ou une adénine désaminase (Goble *et al.*, 2011) qui permet la transition d'une adénine en guanine. Cette technique possède cependant des limites car, outre le fait qu'elle nécessite la présence d'un PAM proche des nucléotides à modifier, toutes les transitions nucléotidiques ne sont pas encore possibles.

Le développement de la technique du « prime editing » permet de dépasser cette dernière limites (Anzalone *et al.*, 2019). En effet, cette nouvelle méthode utilise une nCas9 fusionnée à une reverse transcriptase, fonctionnant en association avec un « prime editing guide RNA » (pegRNA), qui à la fois cible la zone à éditer dans le génome et code pour l'édition que l'on souhaite appliquer (Anzalone *et al.*, 2019).

1.1.3. Etat de l'art des technologies CRISPR-Cas9 chez le pommier

La principale limite à l'emploi des techniques de type CRISPR-Cas9 chez le pommier est la faible efficacité de transformation de cette espèce (Chevreau *et al.*, 2019). A ce jour, seules quatre études ont été publiées. La première mutagenèse ciblée de pommier a été réalisée par Nishitani *et al.* (2016), qui ont édité le gène de la phytoène desaturase (PDS) du pommier avec une efficacité de 32 % chez le porte-greffe de pommier 'JM2'. La même année, Malnoy *et al.* (2016) ont utilisé des ribonucléoprotéines CRISPR-Cas9 pour éditer avec succès des protoplastes de pommier 'Golden Delicious', mais aucune plante éditée n'a été régénérée. Dans notre laboratoire des taux élevés d'extinction par CRISPR-Cas9 de 2 gènes (PDS et TFL1) ont été obtenus chez le pommier 'Gala' (Charrier *et al.*, 2019). Enfin récemment, la combinaison d'une mutagenèse ciblée par CRISPR-Cas9 et d'une élimination du T-DNA par une recombinase inductible a permis l'obtention de mutants de 'Gala' et 'Golden Delicious' moins sensibles au feu bactérien (Pompili *et al.*, 2020).

1.2. Mutagénèse dirigée du gène *MET1*

Dans un premier temps nous avons cherché à obtenir des mutants de pommier pour le gène *MdMET1* afin de valider la fonction de ce gène chez le pommier. Aucun travail de ce type n'a encore été publié chez cette espèce.

Le gène *MET1*, ou *Methyltransferase 1* est impliqué dans la maintenance de la méthylation en contexte symétrique CG (Finnegan *et al.*, 1996, 1998). Sa mutation provoque, chez *A.thaliana*, une diminution globale de la méthylation en contexte CG (Finnegan *et al.*, 1996). Les plantes mutées présentent des phénotypes anormaux, tels que des feuilles et des fleurs modifiées.

Il existe deux orthologues du gène *AtMET1* (AT5G49160) chez le pommier, annotés dans la V1.1 du génome de référence GDDH13 sous les identifiants : MD02G1115100 et MD15G1231100.

1.3. Modification de l'épigénome du promoteur du gène *TFL1*

Dans un second temps nous avons voulu réaliser chez le pommier une preuve de concept de la possibilité d'éditer le niveau de méthylation d'un gène afin de modifier son niveau de transcription. Nous avons choisi pour cela le gène de répression de la floraison *Terminal flowering locus 1* (*Tfl1*, Mimida *et al.* 2009). Ce gène est impliqué dans la répression de la floraison durant la période juvénile d'une plante et sa mutation permet une floraison précoce chez *A. thaliana* (Shannon & Ry Meeks-Wagner, 1991) ou *Pisum sativum L.* (Foucher *et al.*, 2003). Sa répression via la voie des ARNs interférents permet également une floraison précoce chez le pommier (Kotoda *et al.*, 2003).

L'équipe ResPom (IRHS à Angers) a récemment obtenu des mutants éteints pour le gène *MdTFL1* par la méthode CRISPR-Cas9 sur la variété 'Gala' (Charrier *et al.*, 2019), démontrant ainsi qu'une lignée de pommier éteinte pour ce gène est capable de fleurir *in vitro* dans les mois suivants son extinction.

Deux gènes homologues de *Tfl1* ont été identifiés chez le pommier (Mimida *et al.*, 2009) et annoté dans la V1.1 du génome de référence sous les identifiants : MD12G1023900 et MD14G1021100. Pour réaliser l'édition du génome nous avons ciblé le promoteur du gène *MdTFL1* (MD12G1023900). L'idée était de le sur-méthyler afin de réduire son niveau de transcription et ainsi obtenir des plantes fleurissant précocement. L'édition de l'épigénome via une deadCas9 fusionnée à une méthylase est très récente. A notre connaissance, les exemples actuellement disponibles ne concernent que les cellules de mammifère (Liu *et al.*, 2016; McDonald *et al.*, 2016; Vojta *et al.*, 2016; Lei *et al.*, 2017; Xiong *et al.*, 2017). Ces études ont montré une relation entre une hyper-méthylation induite au niveau de la région promotrice et une diminution de la transcription du gène associé. (Liu *et al.*, 2019) ont obtenu des résultats similaires via la déméthylation d'un gène chez *Bombyx mori*, engendrant une augmentation de la transcription de ce gène.

D'autres études ont montré la possibilité d'activer l'expression spécifique d'un gène méthylé chez *A.thaliana* via l'utilisation d'activateur couplé à une Cas9 inactivée (Lowder *et al.*, 2015; Thakore *et al.*, 2015). Papikian *et al.* (2019) ont également montré cette possibilité, toujours chez *A.thaliana* et ils ont constaté que l'activation d'un gène ayant une région promotrice méthylée, avec cette technique, engendre une diminution de la méthylation de la région promotrice en question.

Dans un premier temps nous avons donc cherché à créer une construction basée sur une deadCas9 fusionnée à une méthylase, permettant d'éditer l'épigénome du pommier.

2. Tentative d'extinction d'une voie de maintenance de la méthylation chez le pommier par édition des gènes codant pour la fonction MET1

2.1. Matériel et méthodes

Matériel végétal :

Nous avons utilisé des vitro plants des génotypes GDDH13 et 'Gala'. Les plants initiaux m'ont été fournis par l'équipe ResPom. Ces plants sont cultivés sous jours long à 22°C dans des milieux de culture Murashige and Skoog (1962) optimisés pour la variété 'Gala' comme décrit dans Chevreau et al. (2019).

Clonages

Lors du développement de la technique CRISPR-Cas9, il est nécessaire que les séquences crRNA soient complémentaires des séquences juxtaposées à une séquence nommée PAM (protospacer adjacent motif, **Error! Reference source not found.B**). Des programmes en accès libres tel que CRISPOR-Tefor® (<http://crispor.tefor.net/crispor.py>, Haeussler et al. 2016) permettent d'identifier les crRNAs répondant à cette contrainte et leurs associent un score en fonction de nombreux autres paramètres. Nous avons sélectionné les crRNAs sur le génome de GDDH13, puis nous avons vérifié la présence ou non de mésappariements pour ces crRNAs vis à vis de la variété 'Gala'. Nous avons utilisé le logiciel MultAlin® (<http://multalin.toulouse.inra.fr/multalin/>, Corpet 1988) pour réaliser les alignements entre les séquences des crRNA et celles des gènes ciblés, amplifiées sur l'ADN de 'Gala' et séquencées. Les séquences crRNAs ainsi définies doivent ensuite être clonées dans une cassette d'expression comportant le tracrRNA et un promoteur permettant l'expression de l'ensemble. La synthèse des gRNAs a été commandée à la société IDT (Integrate DNA technologies, Etats Unis d'Amérique, production : Belgique). Chaque gRNA (**Error! Reference source not found.A** et **B**) comporte ainsi des extrémités attB pour le clonage Gateway®, un promoteur U3 ou U6 de pommier (Charrier *et al.*, 2019), un crRNA choisi grâce au logiciel CRISPOR-Tefor®, la séquence du tracrRNA et un site BsaI pour réaliser une ligation Golden Gate (**Error! Reference source not found.C**, Engler, Kandzia, and Marillonnet 2008). Les clonages ont été réalisés comme suit :

1) Golden Gate

Les 500ng de fragments synthétisés sont re-suspendus dans 10µL d'eau ultra pure pour avoir des concentrations de 50ng/µL. Le mix suivant a été réalisé pour la réaction de Golden Gate : 4µL d'H₂O, 1µL de T4 ligase Buffer (10X, Promega, Madison, WI, États-Unis), 1µL de cutsmart buffer (10X, NEB, Ipswich, Massachusetts, États-Unis), 1,5µL des deux fragments à liguer (*n.b.* : les fragments MD15-1 et MD15-2 sont ligués ensemble, et les fragments MD02-1 et MD02-2 sont ligués ensemble), 0,75µL de T4 DNA ligase (Promega) et 0,25µL de BsaI (NEB). Le mix ainsi réalisé est ensuite placé pendant 7h dans un thermocycleur où il subit le programme suivant : (1) 2' à 37°C, (2) 5' à 16°C, (3) retour à l'étape 1 pour 50 cycles, (4) 5' à 50°C, (5) 5' à 80°C.

2) Gateway® (Hartley *et al.*, 2000)

Le produit de la réaction de Golden Gate précédente est utilisé pour réaliser la réaction BP du système de clonage Gateway®. Afin d'intégrer la séquence d'ADN libre dans un plasmide, 10µL de la réaction de Golden Gate sont mélangés à 0,5µL de plasmide pDonrTM207 (à 150ng/µL, obtenu auprès de l'équipe ResPom (IRHS) et 3µL de BP clonase

II enzyme mix (Thermo Fisher Scientific, Waltham, Massachusetts, États-Unis). La réaction est incubée à 25°C sur la nuit. La réaction de recombinaison est stoppée par l'ajout de 1µL de protéinase K (kit BP clonase II enzyme mix) et placée en incubation 10' à 37°C, avant son utilisation à l'étape suivante.

Pour la transformation de bactéries compétentes, la souche d'*E. coli* utilisée est la souche Top10® (Thermo Fisher Scientific). J'utilise 3µL de la réaction de Gateway précédente à laquelle j'ajoute à 50 µl de bactéries compétentes. Je laisse incuber 30' sur glace avant de réaliser un choc thermique de 30'' à 42°C. Les bactéries sont ensuite replacées sur glace pendant 5' avant de recevoir 250µL de milieu SOC (fournit dans le kit) à température ambiante. La suspension bactérienne est ensuite placée en incubation pendant 2h à 37°C et sous agitation. La suspension bactérienne est ensuite étalée sur boîte de Petri contenant du milieu LB (Bertani, 1951) et de l'agarose, additionné d'antibiotique gentamicine (10µg/mL final) et est mise à incuber pour la nuit à 37°C. Certaines des colonies formées sont isolées et leurs plasmides extraits via le kit PureYield™ Plasmid Miniprep System (Promega) suivant les recommandations du fournisseur. L'intégration des fragments et leur intégrité sont vérifiées par PCR puis séquençage (société Genewiz, Radolfzell, Allemagne).

Une seconde étape de recombinaison (étape LR du système de clonage Gateway®) est ensuite réalisée entre le plasmide pDonr™207 et le plasmide de destination pDECas9-Kr (obtenue auprès de Marianne Mazier (INRA Avignon, partenaire dans le projet Genius), pour associer les séquences contenant les crRNA et la cassette d'expression de la Cas9. La sélection de bactéries transformées s'effectue cette fois-ci grâce à l'antibiotique spectinomycine (50µg/mL final). L'intégration des fragments et leur intégrité sont vérifiées par PCR puis séquençage (société Genewiz, Radolfzell, Allemagne), avant le clonage dans la souche d'*Agrobacterium tumefaciens* EHA105 (Hood *et al.*, 1993) portant également le plasmide ternaire pBBR1MCS-5 contenant un gène virG constitutif (van der Fits *et al.*, 2000) (Annexe 4) pour transformation *in planta*.

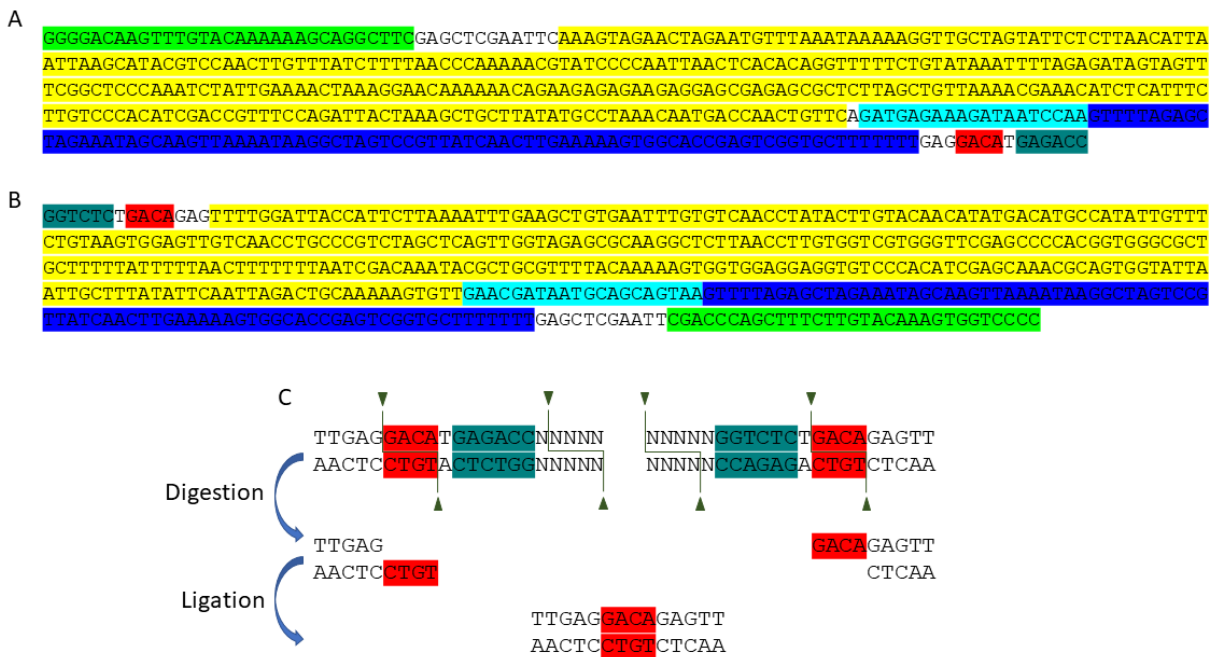


Figure 2.1-1: Motif de restriction de l'enzyme *BsaI* utilisé pour le Golden Gate et construction des gRNA. A. Composition d'un gRNA avec le promoteur U3 du pommier. B. Composition d'un gRNA avec le promoteur U6 du pommier. En Bleu-vert : site de reconnaissance de l'enzyme *BsaI* ; en rouge: séquence choisie pour la ligation Golden Gate ; en vert clair : extrémités attB pour la réaction BP du clonage Gateway ; en jaune : promoteur U3 ou U6 ; en bleu clair : crRNA ; en bleu foncé : tracrRNA. C. Principe du golden Gate reposant sur l'enzyme de restriction *BsaI*, appliqué aux gRNA présentés en A et B. L'enzyme *BsaI* reconnaît la séquence en bleu-vert et coupe comme indiqué par les flèches. Les séquences en rouge ont été choisies et ajoutées de telle manière qu'après coupure, elles donnent des extrémités cohésives pouvant se liguer entre elles. A noter qu'une fois la réaction de Golden Gate réalisée, le site *BsaI* n'est plus présent dans la séquence.

Transformation

La transformation de feuilles de vitro-plants de pommier est réalisée comme décrit par Chevreau et al. (2019). Brièvement, une suspension d'agro-bactéries EHA105 contenant la construction d'intérêt est produite à une concentration ajustée à 10⁸ bactéries par mL. Du Silwet, un agent surfactant, est ajouté à 0,002% pour améliorer l'efficacité de pénétration des bactéries dans les espaces intercellulaires des feuilles durant l'infiltration sous vide. Dans un flacon stérile de 50mL, 20 à 30mL de la suspension d'agro-bactéries sont mises en contact avec 30 à 50 jeunes feuilles de pommier issues de plant *in vitro* présentant un bon état physiologique. Entre 300 et 400 feuilles sont utilisées au total pour une transformation. Une infiltration sous vide est ensuite réalisée pendant 1 min. Les feuilles sont ensuite récupérées et scarifiées (4 à 5 coupures par feuille) puis placées, face abaxiale au-dessus, sur un milieu de coculture sans sélection pour 2 jours. Au bout des deux jours les feuilles sont transférées sur un milieu sélectif (céfotaxime (300mg/L), timentin (150mg/L) et kanamycine (100mg/L)) et placées à l'obscurité à 22°C.

Tous les mois pendant 6 mois, les feuilles sont transférées sur un nouveau milieu de culture sélectif. Lors de l'apparition de bourgeons, les feuilles portant des bourgeons sont placées à la lumière pour que les bourgeons se développent. Une fois bien développés, les bourgeons sont placés sur un milieu sélectif ne contenant aucune hormone, pendant un mois. A cette étape chaque bourgeon est identifié par un code (numéro de la transformation et numérotation du bourgeon). Les bourgeons

sont ensuite placés sur un milieu de culture sélectif de type POM (Annexe 5, Murashige and Skoog (1962)).

Pour les différentes transformations, j'ai également utilisé une souche EHA105 portant une construction 35S::GUS::GFP comme témoin de transformation. Cette construction m'a été fournie par l'équipe ResPom.

Vérification de l'état transgénique des lignées et séquençage des zones ciblées par le système CRISPR-Cas9

Une fois les pommiers transgéniques suffisamment développés pour être prélevés, j'ai extrait leur ADN en utilisant le Kit d'extraction NucleoSpin Plant II kit (Macherey-Nagel, Hoerd, France). J'ai dans un premier temps vérifié que ces lignées étaient bien transgéniques avec différents couples d'amorces. Les amorces utilisées avaient pour but 1) de tester la bonne qualité de l'ADN extrait via l'amplification d'un gène de ménage (*Efl α* , amorces n° 9 et 10, annexe 6) ; 2) de tester l'absence de contamination des lignées par des agro-bactéries (amorces n° 13 et 14, annexe 6), car si tel est le cas on ne peut pas faire confiance aux amplifications suivantes, qui peuvent aussi se faire à partir des plasmides portés par les agro-bactéries ; 3) de tester la présence de la résistance à la kanamycine (NPTII, amorces n°15 et 16, annexe6) et enfin 4) de tester la présence des cassettes Cas9 (amorces n°21 et 22, annexe 6) + crRNA (amorces n°19 et 20, annexe 6). L'ADN a ensuite été amplifié par PCR, avec des amorces (n°1, 2, 3 et 4, annexe 6) dessinées de part et d'autre des sites ciblés par la Cas9 via les crRNA. Ces PCR peuvent ensuite être envoyées directement à séquencer (Genewiz, Radolfzell, Allemagne) ou bien servir de matrice pour un clonage en masse à l'aide du kit pGEM®T Easy (Promega), d'après les recommandations du fournisseur. L'idée étant de séparer les éventuels différents produits présents dans le mélange PCR. L'ADN de plusieurs clones bactériens, portant chacun l'un des produits du mélange est ensuite envoyé à séquencer.

Acclimatation et phénotypage en serre

Les plants *in vitro* ont été enracinés pour leur permettre une croissance en serre, suivant le protocole décrit par Faize et al. (2003). Brièvement, les plants sont placés pendant sept jours à l'obscurité sur un milieu d'induction racinaire riche en auxine, puis transférés sur un milieu d'expression racinaire sans hormones à la lumière pendant au minimum trois semaines. Les plants sont ensuite repiqués en serre, en mottes fertilisées et sous couvert pour conserver une forte humidité. Le couvert est progressivement retiré pour permettre l'acclimatation progressive des plants. Des observations macroscopiques ont été réalisées à différents stades de développement des plants en serre.

Epigénotypage

→ Restriction enzymatique par *HpaII*.

L'ADN de feuilles de plants enracinés et mis en culture en serre a été extrait suivant le protocole développé par Fulton, Chunwongse, and Tanksley (1995). Trois extractions sont réalisées par plante à partir de prélèvements de feuilles différentes. 5 μ g d'ADN ont été utilisés pour la digestion enzymatique dans le mix suivant : 20 μ L de tampon, 12 μ L d'enzyme et de l'eau ultra pure pour compléter le volume à 200 μ L. La digestion a été réalisée à 37°C pendant une nuit puis l'enzyme est inactivée à 80°C pendant 20'. J'ai ensuite ajouté 1 μ L de glycérol bleu pour marquer les culots d'ADN, 200 μ L d'eau ultra pure, 40 μ L de NaAc 3M et 1mL d'éthanol pur pour réaliser une précipitation à -80°C pendant 2h. Les culots sont ensuite lavés et re-suspendus dans 50 μ L d'eau ultra pure. La concentration a été mesurée au nanodrop et ajustée à 2ng. μ L⁻¹ pour réaliser les PCR. Des témoins non digérés sont réalisés en suivant les mêmes étapes mais avec ajout d'eau en

remplacement de l'enzyme *HpaII*. A la suite, des PCRs sont réalisées en utilisant les amorces pHODOR (taille attendue de l'amplicon : 283pb, annexe 6) ainsi que des amorces pour les gènes de ménage EF1 α (témoin pour la quantité de matrice initiale, ne possédant pas de site *HpaII*, aucune coupure n'est donc attendue, taille attendue de l'amplicon : 400pb). Afin d'observer des différences de quantité sur gel d'électrophorèse (intercalant : Midori Green, nippon genetics), les PCRs sont réalisées en réduisant le nombre de cycle d'amplification à 25 cycles (au lieu de 35). L'analyse de l'intensité des bandes sur gel d'agarose est réalisée avec le logiciel ImageJ®. L'intensité de chaque bande pour chaque amplification est mesurée. Une normalisation des intensités mesurées sur l'amplification HODOR est réalisée en les divisant par l'intensité mesurée pour l'amplification du gène de ménage EF1 α du même échantillon. Pour chaque génotype et chaque répétition, je réalise ensuite un ratio avec les valeurs normalisées des amplifications de HODOR, en divisant la valeur obtenue pour l'échantillon non digéré par la valeur obtenue pour l'échantillon digéré. Les répliques ne présentant pas de diminution d'intensité entre les échantillons non digéré et digéré (ratio inférieur à 1) ne sont pas pris en compte par la suite. Je réalise ensuite une moyenne des valeurs obtenues pour les répliques pour un même génotype.

Séquençage bisulfite.

L'ADN a été extrait par kit (Macherey-Nagel) puis envoyé pour création des bibliothèques et séquençage bisulfite du génome complet sur la plateforme Génome Québec (<http://www.genomequebec.com>).

2.2. Résultats

2.2.1. Recherche des orthologues de *AtMet1* chez le pommier

a) V1.0 du génome de GDDH13

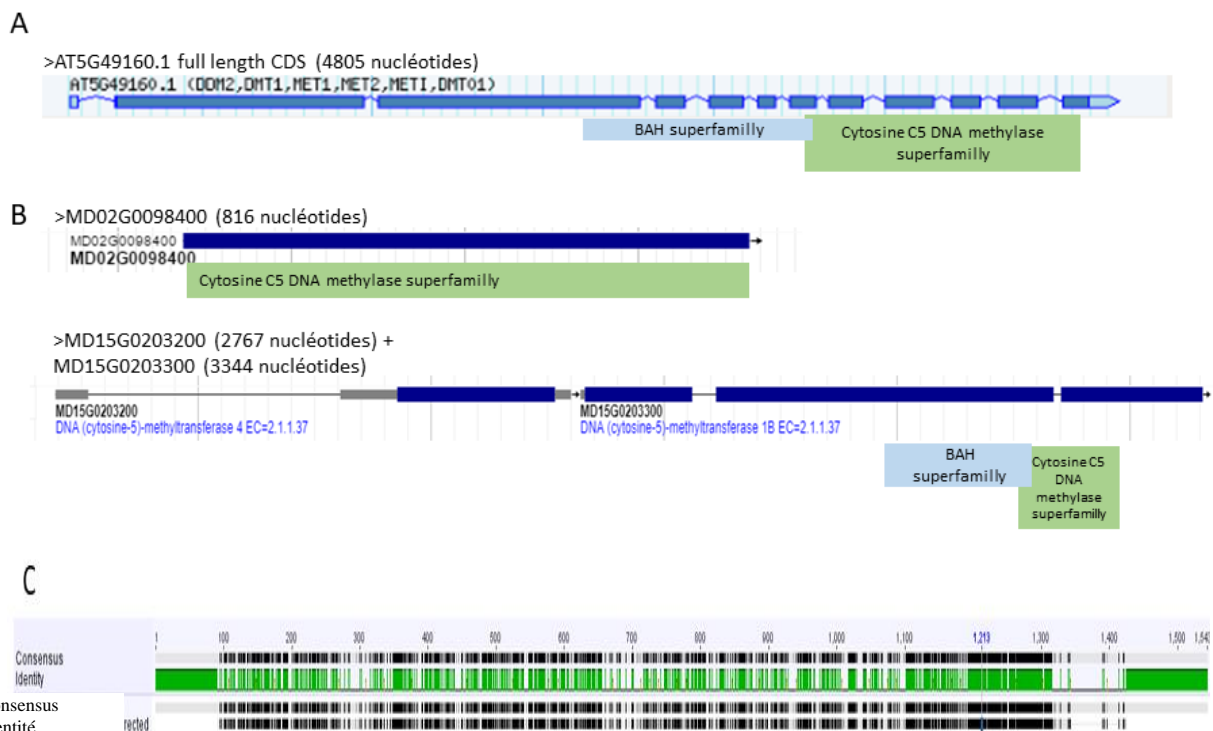
Au début de ce projet j'ai commencé à travailler avec la V1.0 du génome de GDDH13 (Daccord *et al.*, 2017). Dans un premier temps j'ai recherché les gènes annotés comme étant des méthyltransférases et j'en ai identifié 31 (Annexe 7). Seul un de ces gènes était annoté comme orthologue à *AtMET1* : MD02G0098400. Comme le génome du pommier est dupliqué (Daccord *et al.*, 2017) je m'attendais à trouver un deuxième gène pour la fonction MET1. J'ai sélectionné le gène MD15G0203300 qui est annoté comme homologue à *AtMET2* d'après l'annotation faite sous TAIR (**Error! Reference source not found.**A, TAIR 2020).

J'ai ensuite observé ces séquences sur un JBrowse local (**Error! Reference source not found.**B) et je me suis aperçu de deux choses. D'abord le gène MD02G0098400 (*Md02MET1*) n'était probablement pas complet : en amont du gène il y avait une zone de « N », c'est-à-dire une zone non assemblée/séquencée entre deux contigs. De plus la séquence nucléotidique du gène *Md02MET1* était de 816 nucléotides contre 4805 nucléotides pour la séquence d'*AtMET1*.

Ensuite pour le gène MD15G0203300 (*Md15MET1*), j'ai identifié un second gène annoté comme une méthyltransférase à proximité, MD15G0203200 (**Error! Reference source not found.**B), que je n'avais pas identifié lors de la recherche initiale. Du fait de leur position relativement proche j'ai pensé que ces deux gènes ne formaient probablement qu'un unique gène. J'ai donc récupéré la séquence codante annotée pour chacun des deux gènes et j'ai découvert sur le JBrowse (**Error! Reference source not found.**B) que les deux gènes en question étaient séparés par 47 nucléotides.

J'ai donc rajouté ces nucléotides entre les séquences codantes des deux gènes. J'ai ensuite aligné cette séquence (de 6156 nucléotides) corrigée sur le gène *AtMET1*, et j'ai trouvé une homologie de séquence de 68% au niveau nucléotidique et de 58% au niveau de la séquence en acides aminés (**Error! Reference source not found.C**).

J'ai ensuite observé l'annotation des domaines fonctionnels de *AtMET1* et regardé si je retrouvais les mêmes domaines dans les séquences protéiques des gènes annotés *MET1* chez le pommier. Pour la séquence *AtMET1* (**Error! Reference source not found.A**) j'ai identifié les domaines DNMT1-RFD. Ils permettraient un ciblage du complexe de réplication par la méthyltransférase afin de restaurer la méthylation à la suite de la réplication. Les domaines de la famille des BAH (bromo-adjacent homology) qui auraient un rôle dans le repliement et l'adressage des méthyltransférases au niveau de la fourche de réplication, afin d'assurer la maintenance de la méthylation lors de la duplication de l'ADN (Yarychkivska *et al.*, 2018). Et un domaine DMC (DNA-cytosine methylase) appartenant à la famille des méthylases spécifiques de la méthylation des cytosines en C5 (méthylation de l'ADN). Pour la protéine MdMET1 issue du gène *Md02MET1*, le résultat du Blastx indique une homologie de séquence avec le domaine DMC uniquement. Pour la protéine MdMET1 du gène *Md15MET1* + MD15G0203200 j'observe les mêmes domaines que ceux présents dans *AtMET1* avec une disposition similaire. Ceci confirmait également que le gène *Md02MET1* (MD02G0098400) était vraisemblablement incomplet dans la version 1.0.



Structures géniques et protéiques des gènes codant pour la fonction MET1 chez *A. thaliana* et *M. domestica*. A. structure du gène chez *A. thaliana* (<https://www.arabidopsis.org>) et localisation des régions codantes pour les fonctions associées à la méthylation de l'ADN. B. structure des gènes chez *M. domestica* dans la V1.0 du génome de GDDH13. C. alignement des séquences protéiques de AT5G49160 MD15G0203200+MD15G0203300.

b) V1.1 du génome de GDDH13

Une fois la V1.1 du génome de GDDH13 disponible j'ai recherché les gènes *MET1* déjà identifiés. La numérotation des gènes ayant été modifiée, les deux gènes *MET1* identifiés précédemment ont maintenant les identifiants MD02G1115100 (*Md02MET1*) et MD15G1231100 (*Md15MET1*). Le gène situé sur le chromosome deux présentait également un nouvel assemblage. Sa structure étant maintenant similaire à celle de son homologue sur le chromosome 15. En comparant la séquence codante des gènes (**Error! Reference source not found.**A) j'ai trouvé que leurs séquences nucléotidiques présentaient un pourcentage d'identité de 94%. Pour la séquence protéique, la similarité entre les deux gènes était de 92%.

Ces deux gènes sont présents au sein des régions dupliquées entre les chromosomes 2 et 15 de GDDH13 (Daccord *et al.*, 2017). Ils sont probablement issus du dernier évènement de duplication du génome (Daccord *et al.*, 2017).

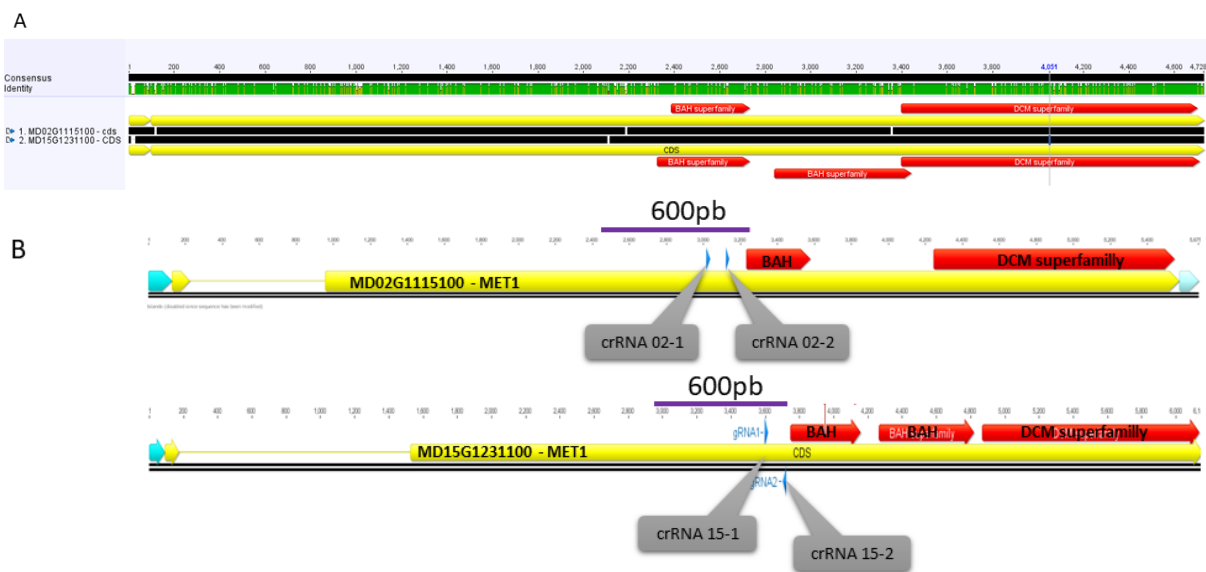


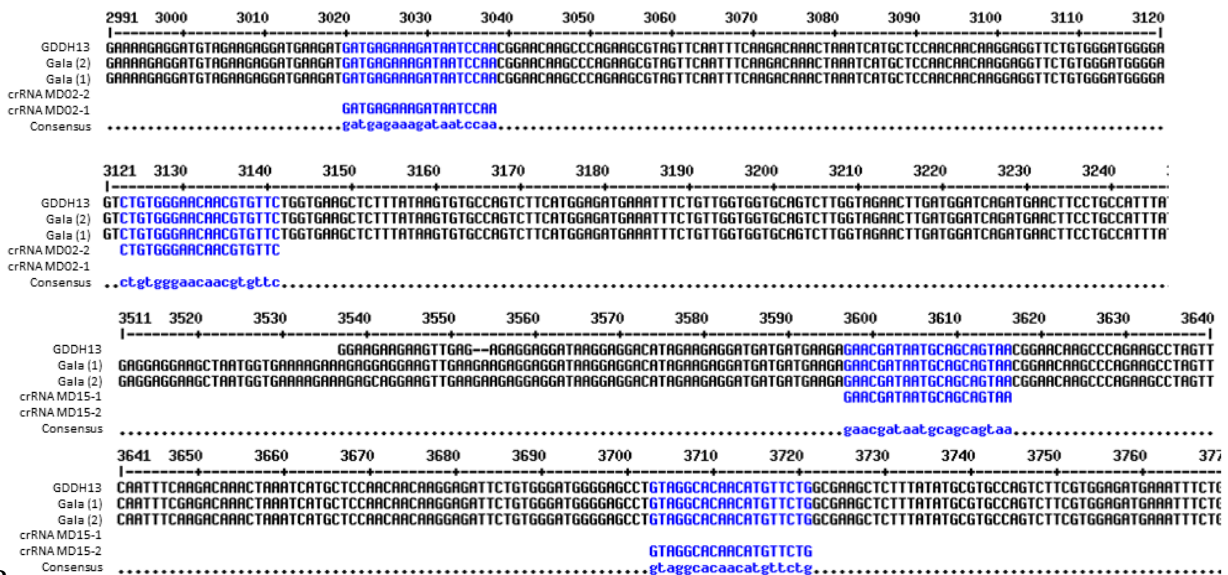
Figure 2.2-2: Comparaison de la séquence des gènes codants pour la fonction MET1 sur la V1.1 du génome de référence GDDH13. A. Alignement de la séquence codante des gènes MET1 (séquence nucléotidique). B. Localisation des crRNAs choisis sur les séquences MD02G1115100 et MD15G1231100 . Le trait violet indique la localisation de la séquence utilisée pour identifier les crRNAs à l'aide du site CRISPOR-Tefor (<http://crispor.tefor.net>).

c) Chez le génotype 'Gala'

Pour l'équipe ResPom, la variété 'Gala' est la variété de référence pour la transformation du pommier. Afin d'optimiser la probabilité d'obtenir des mutants, nous avons donc réalisé les transformations sur le génotype GDDH13 et sur le génotype 'Gala'. Le génome de 'Gala' a récemment été séquencé par une équipe de l'Agroscope en Suisse ('communication personnelle'). Nous leur avons transmis les séquences de *MET1* de GDDH13 afin qu'ils recherchent les séquences homologues sur ce génome. Les alignements des deux séquences *Md02MET1* et *Md15MET1* n'ont fait ressortir qu'une seule séquence, correspondant au gène *Md15MET1*. Le gène *Md02MET1* n'a donc pas été retrouvé dans leur génome.

En réalisant des PCRs et séquençages à l'aide d'amorces spécifiques (amorces N°1 à 4, annexe 6) à *Md02MET1* ou *Md15MET1*, dessinées sur la séquence de GDDH13, j'ai ultérieurement pu confirmer la présence du gène *Md02MET1* chez 'Gala' (**Error! Reference source not found.A**).

A.



B.

Figure 2.2-3: Alignement entre les crRNA dessinés sur le génome de référence GDDH13 et les résultats de séquençage de PCR faites sur de l'ADN extrait de 'Gala'. A. alignement entre les séquençages en sens (1) et antisens (2) de l'amplification PCR réalisée avec les amorces N°1 et 2 (annexe 6) sur le gène MD02G1115100 de 'Gala' avec le génome de référence GDDH13 V1.1. B. alignement entre les séquençages en sens (1) et antisens (2) de l'amplification PCR réalisée avec les amorces N°3 et 4 (annexe 6) sur le gène MD15G1231100 de 'Gala' avec le génome de référence GDDH13 V1.1. Sont également positionnés les crRNA MD02-1 et MD02-2 dessinés sur le génome de référence GDDH13 V1.1. B. alignement entre les séquençages en sens (1) et antisens (2) de l'amplification PCR réalisée avec les amorces N°3 et 4 (annexe 6) sur le gène MD15G1231100 de 'Gala' avec le génome de référence GDDH13 V1.1. Sont également positionnés les crRNA MD15-1 et MD15-2 dessinés sur le génome de référence GDDH13 V1.1. Alignements réalisés avec le logiciel MultAlin (<http://multalin.toulouse.inra.fr/multalin/>).

2.2.2. Mise en place de la technique CRISPR-Cas9 pour réaliser l'extinction des gènes *Md02MET1* et *Md15MET1* du pommier

Après discussions avec mes encadrants et mon comité de suivi de thèse, nous avons estimé qu'il était probable que la mutation simultanée des deux gènes *MET1* soit délétère, nous avons donc décidé de réaliser dans un premier temps des simples mutants pour chaque gène *MET1* et un double mutant dans un second temps.

a) Sélection des crRNAs pour le génotype GDDH13

Pour réaliser un gRNA spécifique de chaque gène, j'avais deux contraintes : trouver des sites PAM en amont du domaine prédit portant l'activité méthyltransférase et situés à proximité d'une séquence pouvant servir de crRNA, cette dernière devant être divergente entre les deux gènes *MdMET1* (**Error! Reference source not found.**B). Selon ces critères et pour chaque gène, j'ai sélectionné deux crRNAs (**Error! Reference source not found.**B, Annexe 8), que j'ai espacés d'une centaine de paires de bases afin d'optimiser la probabilité d'obtenir une grande délétion (Zhao *et al.*, 2016; Do *et al.*, 2019). L'utilisation de deux crRNAs permet également une augmentation de l'efficacité de mutation, car les crRNAs n'ont pas toujours la même efficacité (Do *et al.*, 2019).

Afin de détecter les crRNAs dans la région que je souhaitais cibler, j'ai soumis 600pb de chaque gène (séquence de GDDH13) au logiciel CRISPOR-Tefor®. Ces 600 pb sont situées en amont des sites BAH de chaque gène (**Error! Reference source not found.**B).

Le logiciel CRISPOR permet d'obtenir différentes informations concernant la qualité des crRNA (**Error! Reference source not found.**, Annexe 8 tableaux 1 et 3). Les informations que j'ai considérées sont : l'absence de site de restriction utilisé dans la procédure de clonage ultérieure et un score de spécificité élevée. Du fait de l'homologie de séquence entre les deux gènes, de nombreux crRNAs proposés étaient identiques entre les deux gènes et ont donc été éliminés. J'ai finalement identifié deux séquences pour chaque gène (**Error! Reference source not found.**B).

Une autre donnée importante est la présence de séquences alternatives d'appariement (« off target », Annexe 8 tableau 2 et 4). Il est important d'avoir un nombre de séquences alternatives d'appariement le plus faible possible, d'autant plus si les séquences présentent moins de deux mésappariements avec le crRNA. Charrier *et al.* (2019) ont en effet montré que dès trois mésappariements, les séquences alternatives ne semblent plus coupées par l'enzyme. Pour les crRNAs sélectionnés sur le gène MD15G1231100 par exemple (Annexe 8 tableau 2) il y a 10 séquences alternatives d'appariement pour le crRNA 15-1 et 33 pour le crRNA 15-2. Pour chaque séquence alternative d'appariement, il y a au moins deux mésappariements avec le crRNA considéré. La quasi-totalité des séquences alternatives d'appariements est située soit dans des régions intergéniques soit dans des introns. Les séquences alternatives d'appariement situées dans des exons présentent un minimum de 4 mésappariements. On note également qu'aucune séquence alternative d'appariement n'est située sur le gène MD02G1115100.

	Position/ Strand	Guide Sequence + PAM + Restriction Enzymes <input type="checkbox"/> Only G- <input type="checkbox"/> Only GG- <input type="checkbox"/> Only A-	MIT Specificity Score	CFD Spec. score	Predicted Efficiency <small>Show all scores</small> Doench '16 Mor-Mates	Outcome Out-of-Frame Lindel	Off-targets for 0-1-2-3-4 mismatches + next to PAM	Genome Browser links to matches sorted by CFI <input type="checkbox"/> exons only <input type="checkbox"/> Chr15 only	
crRNA-15-1	470 / fw	GAACGATAATGCACAGTAA CCG Enzymes: <i>MaeIII, TspGI</i> Cloning / PCR primers	98	99	53	42	66 76	0-0-1-0-9 0-0-0-0-0 10 off-targets	4:intron:mRNA:MD01G1048700 4:intergenic:mRNA:MD06G1120300-mRNA:MD06G1120400 4:exon:mRNA:MD06G1006700 show all...
crRNA-15-2	553 / rev	CCAGAACATGTTGTGCCTAC AGG Enzymes: <i>LmnI, BlnI, NlaIV</i> Cloning / PCR primers	89	95	51	49	77 83	0-0-1-5-27 0-0-0-0-0 33 off-targets	3:intergenic:mRNA:MD06G1074900-mRNA:MD06G1075100 2:intergenic:mRNA:MD14G1163300-mRNA:MD14G1163400 3:intergenic:mRNA:MD16G1178900-mRNA:MD16G1179200 show all...
crRNA-02-1	272 / fw	TGATGAGAAAGATAATCCAA CCG Enzymes: <i>MmeI, TspGI</i> Cloning / PCR primers	84	88	70	46	63 79	0-0-1-14-104 0-0-0-0-0 119 off-targets	4:intron:mRNA:MD12G1087600 2:intergenic:mRNA:MD16G1170700-mRNA:MD16G1170800 3:intron:mRNA:MD03G1009000 show all...
crRNA-02-2	375 / fw	TCTGTGGGAACAAGTGTTC TGG <small>Inefficient</small> Enzymes: <i>AsuHPI, AflIII</i> Cloning / PCR primers	96	98	43	33	65 80	0-0-1-0-13 0-0-0-0-0 14 off-targets	2:intergenic:mRNA:MD05G1230600-mRNA:MD05G1230700 4:exon:mRNA:MD15G1231100 4:intergenic:mRNA:MD02G1175500-mRNA:MD02G1175700 show all...

Figure 2.2-4 : Sélection des crRNAs sur le logiciel CRISPOR-Tefor. Capture d'écran des différents crRNAs sélectionnés et de leurs caractéristiques de qualités. Code couleur: vert pour les crRNA conseillé, jaune/orange pour les crRNA de moyenne qualité, rouge pour les crRNA déconseillés. « Position/strand » : position relative de la séquence par rapport à la séquence de 600pb utilisée pour le dessin. « Guide sequence + PAM » : séquence du crRNA et du PAM, cette colonne présente également les sites de restriction provoquant un clivage de la séquence. « MIT and CFD specificity score » : spécificité du crRNA, plus le score est élevé (0-100) moins le crRNA risque de s'hybrider sur des séquences alternatives. « Predicted Efficiency » : efficacité relative des crRNAs (de 1 à 100). « Off-target » et « mismatches » : nombre de séquence d'appariement alternatif potentiel du crRNA dans le génome et nombre de mésappariement entre le crRNA et ces séquences (voir tableau 2 et 4 de l'annexe 8 pour les séquences alternatives d'appariement).

b) Sélection des crRNAs pour le génotype 'Gala'

J'ai réalisé un alignement des crRNAs sur les séquences du génotype 'Gala' que j'avais à ma disposition, pour voir si les crRNAs définis pour GDDH13 étaient utilisables aussi pour ce génotype. Au début de la sélection, je n'avais, pour ce génotype, que la séquence du gène *Md15MET1* à ma disposition. L'alignement des séquences ne montrait aucun mésappariement entre les crRNAs du gène *Md15MET1* choisis sur la séquence de GDDH13 et la séquence de 'Gala' (**Error! Reference source not found.B**). Je pouvais donc utiliser les mêmes crRNAs chez ces deux génotypes, pour le gène *Md15MET1*.

J'ai ensuite réalisé des PCRs sur de l'ADN de 'Gala' avec le couple d'amorces dessinées sur GDDH13 pour le gène MD02G1115100, que j'ai ensuite fait séquencer. Comme présenté dans la **Error! Reference source not found.A**, il n'y a pas de mésappariement entre les crRNAs dessinés sur le génome GDDH13 et la séquence du gène MD02G1115100 de 'Gala'. Les crRNA MD02-1 et MD02-2 sont donc utilisables pour cibler le gène *Md02MET1* chez 'Gala' (**Error! Reference source not found.B**)

2.2.3. Résultats des transformations de vitroplants de pommiers

J'ai réalisé huit transformations (hors témoins) en vue d'obtenir des mutants *met1* chez le pommier (**Error! Reference source not found.**). Quatre de ces transformations ont été réalisées sur le génotype 'Gala'. Les transformations N°2 et 8 sont des répliques réalisés pour augmenter l'effectif des mutants. La transformation N°13 est un essai pour cibler le gène *Md02MET1* de 'Gala', non

retrouvé par blast sur l'assemblage du génome de 'Gala', mais retrouvé par PCR. La transformation N°14 est un essai pour obtenir des doubles mutants.

Pour les transformations sur le génotype 'Gala', seules les transformations N°2 et 8 ont permis d'obtenir des lignées transformées et éditées pour le gène *Md15MET1*.

Chaque lignée transformée obtenue est numérotée en fonction de la transformation et son rang lors de l'isolement, comme présenté dans le **Error! Reference source not found.**. Ainsi un bourgeon issu de la transformation numéro 2 et étant le 4^{ème} à être isolé pour cette transformation portera le nom '2D'. Si plusieurs bourgeons sont isolés sur le même cal ou la même feuille, on rajoute une numérotation supplémentaire : ainsi les lignées 2C1 à 2C4 sont issues d'un même cal ou d'une même feuille.

J'ai également réalisé deux transformations pour chacun des gènes *Md02MET1* et *Md15MET1* chez le génotype GDDH13. Seules les transformations N°4 et 5 ont permis d'obtenir des bourgeons, mais nous n'avons pas réussi à les maintenir et les faire se développer.

Les transformations N°9 et 10 sur le génotype GDDH13 ont été un échec, tout comme les transformations n° 13 et 14 sur 'Gala'.

Tableau 2.2-1 : Résumé des transformations réalisées

N° de la transformation	Génotype	Date	Identifiant du (des) gène(s) ciblé(s)	Nombre de bourgeons produits	Nombre de lignées obtenues	Taux de transformation
2	Gala	09.17	MD15G1231100	53	10	2.5%
4	GDDH13	10.17	MD02G1115100	29	0	/
5	GDDH13	10.17	MD15G1231100	25	0	/
8	Gala	11.17	MD15G1231100	15	1	0.25%
9	GDDH13	11.17	MD02G1115100	1	0	/
10	GDDH13	11.17	MD15G1231100	0	0	/
13	Gala	04.19	MD02G1115100	0	0	/
14	Gala	04.19	MD02G1115100 - MD15G1231100	0	0	/
3, 6, 12	Gala ou GDDH13	09.17 10.17 11.17	Construction 35S::GUS::GFP	4, 0, 0	0	/

Tableau 2.2-2 : Bourgeons obtenus dans les transformations 2 et 8 (cf: table 2.2-1), observation du nombre de bande sur gel d'électrophorèse à la suite de l'amplification par PCR de la zone ciblée pour la mutation du gène MD15G1231100.

Transformation n°	Numéro d'isolement	Numéro si plusieurs bourgeons sur le même cal	Nom de la lignée	Nombre de bandes sur gel
2	C	1	2C1	Deux bandes
2	C	2	2C2	Une bande
2	C	3	2C3	Une bande
2	C	4	2C4	Deux bandes
2	D	/	2D	Une bande
2	H	1	2H1	Une bande
2	H	2	2H2	Deux bandes
2	H	3	2H3	Une bande
2	L	/	2L	Deux bandes

2	P	/	2P	Deux bandes
8	A	/	8A	Deux bandes

2.2.4. Analyse moléculaire des lignées transformées

J'ai généré des lignées potentiellement éditées pour le gène *Md15MET1* chez le génotype 'Gala'. Dans un premier temps j'ai validé le caractère transgénique des lignées par PCR. J'ai ainsi pu mettre en évidence que l'ADN est de bonne qualité pour la quasi-totalité des lignées et du témoin non transformé (**Error! Reference source not found.A**). En effet pour les lignées 2H2, 2L et 8A je n'ai pas eu d'amplification pour le gène de ménage *EF1 α* . Par ailleurs les lignées ne sont pas contaminées par des agro-bactéries, en effet je n'ai pas d'amplification pour le couple d'amorces correspondant, que l'on sait par ailleurs fonctionner (**Error! Reference source not found.A** et **B**). Enfin toutes les lignées ont bien une insertion de la Cas9, des crRNA et du gène de résistance à la kanamycine (**Error! Reference source not found.A**). Pour la lignée 2D l'amplification du gène de résistance à la kanamycine a été peu efficace, mais une bande à la taille attendue est bien observable. Enfin, sur ce premier gel, seule une partie des lignées présentent une amplification pour le gène *Md15MET1* et il semble que certains fragments, dans les lignées potentiellement mutantes, soient d'une taille inférieure à celle du fragment sauvage (**Error! Reference source not found.A**). Une seconde amplification, identique à la précédente, du gène *Md15MET1* dans les différentes lignées (**Error! Reference source not found.C**) montre une amplification d'un fragment de 342 pb pour le témoin 'Gala' non transformé (T-). J'observe ensuite trois profils différents. Soit une bande unique à une hauteur similaire au témoin (342pb, lignées transgéniques 2D, 2H1, 2H3), soit une bande unique à une taille inférieure (lignées transgéniques 2C2, 2C3) soit deux bandes, l'une de taille attendue et l'autre de taille inférieure (lignées transgéniques 2C1, 2C4, 2H2, 2L, 2P, 8A). J'ai ensuite envoyé à séquencer les produits PCR obtenus avec les amorces pour le gène *Md15MET1* pour les 5 lignées qui ne présentaient qu'un seul produit de PCR.

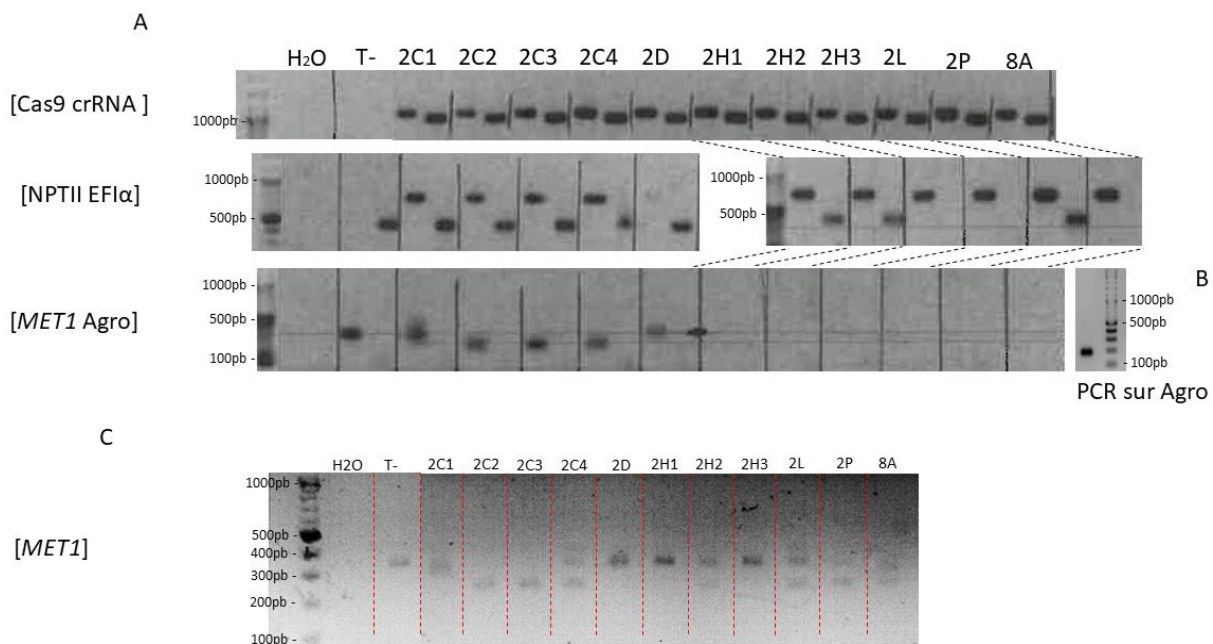


Figure 2.2-5 : Vérification par PCR des lignées potentiellement mutantes pour le gène *MET1* dans le génotype Gala. A. PCR de vérification de l'état transgénique des lignées. Entre crochets les différents couples d'amorces utilisés pour chaque ligne du gel, en colonne les lignées. B. témoin d'amplification du couple d'amorces n°13 et 14, ciblant un gène d'Agrobacterium. C. Seconde PCR réalisée avec les amorces

N°3 et 4 pour amplifier la région potentiellement mutée du gène MD15G1231100. J'ai modifié les contrastes de la photographie pour une meilleure lisibilité. H2O: témoin eau, mix PCR sans matrice ADN, (T-): Gala non transformé.

Sur les 5 lignées séquencées, 3 présentaient un profil de séquençage lisible : les lignées 2C2, 2C3 et 2H1 (**Error! Reference source not found.**). Ces lignées présentent un profil de mutation homozygote c'est-à-dire que les deux allèles présentent la même mutation. Pour les deux autres lignées, le séquençage n'était pas lisible, probablement dû à la production d'un mélange d'amplicons de plusieurs allèles. Tout comme les lignées présentant deux bandes sur la **Error! Reference source not found.C**, ces lignées sont donc soit hétérozygotes pour l'édition (1 seul allèle muté), soit bi-alléliques (des mutations différentes sur chaque allèle) voire chimériques. C'est-à-dire que le profil d'édition (homozygote, hétérozygote ou bi-allélique) du gène *MET1* est peut-être différent selon les cellules. Afin de résoudre les profils de ces lignées, il faudrait réaliser un clonage en masse du produit PCR pour isoler les différents amplicons les uns des autres, puis séquencer plusieurs clones bactériens pour avoir une idée des profils en présence. Puisque nous avons déjà obtenu 3 lignées homozygotes pour la mutation du gène *met1*, nous avons choisi de n'effectuer ce travail que pour une autre plante : 2P.

À la suite de l'obtention de ces mutants, nous avons suspendu la culture des feuilles transformées car nous avons suffisamment de matériel pour l'objectif fixé. Nous aurions sans doute pu obtenir un plus grand nombre de bourgeons en prolongeant la culture.

Pour la suite de mon analyse je ne m'intéresserai qu'aux lignées dont le profil d'édition est connu : les lignées 2C2, 2C3, 2H1 et 2P.

Les lignées 2C2 et 2C3 présentent la même édition (**Error! Reference source not found.A**), à savoir une délétion complète du fragment situé entre les deux crRNA. L'analyse de la séquence protéique de ces mutants (**Error! Reference source not found.B**) montre une modification des acides aminés au niveau du site de coupure du crRNA 15-1. Cette modification se poursuit sur 18 acides aminés avant l'apparition d'un codon Stop. L'édition a donc également provoqué une modification du cadre de lecture.

Pour la lignée 2H1, on observe que pour le crRNA 15-1 la réparation de la coupure a engendré l'insertion d'une thymine, et pour le crRNA 15-2 la coupure a engendré la délétion d'une guanine (**Error! Reference source not found.A**). L'observation de la séquence protéique de 2H1 (**Error! Reference source not found.B**) montre une modification de la séquence des acides aminés, par rapport à la séquence de 'Gala' non transformée. En effet, sept acides aminés suivant la mutation, il y a apparition d'un codon Stop.

Enfin pour la lignée, 2P (**Error! Reference source not found.A**), l'un des allèles de ce mutant ne présente pas de mutation, tandis que le second allèle présente une large délétion similaire à ce qui est observé pour les lignées mutantes 2C2 et 2C3. Nous avons donc un mutant hétérozygote.

A

	2144	
GDDH13		ATGATGAAGA GAACGATAATGCAGCAG-TAA CGGAACAAGCCCGAAGCCTAGTTCAATTTTCGAGACAACTAA
Gala		ATGATGAAGA GAACGATAATGCAGCAG-TAA CGGAACAAGCCCGAAGCCTAGTTCAATTTTCAGACAACTAA
2C2		ATGATGAAGAAGAACGATAATGCAGCAG-----
2C3		ATGATGAAGAGAACGATAATGCAGCAG-----
2H1		ATGATGAAGAGAACGATAATGCAGCAG TTAACGGAACAAGCCCGAAGCCTAGTTCAATTTTCAGACAACTAA
2P - Clone1		ATGATGAAGAGAACGATAATGCAGCAG-----
2P - Clone2		ATGATGAAGAGAACGATAATGCAGCAG-TAACGGAACAAGCCCGAAGCCTAGTTCAATTTTCAGACAACTAA
MD15G1231100		ATCATGCTCCAACAACAAGGAGATTCTGTGGGATGGGGAGCCTGTAGGCACAACATGTTCTGGCGAAGCTCTT
Gala		ATCATGCTCCAACAACAAGGAGATTCTGTGGGATGGGGAGCCTGTAGGCACAACATGTTCTGGCGAAGCTCTT
2C2		-----GGCACAAACATGTTCTGGCGAAGCTCTT
2C3		-----GGCACAAACATGTTCTGGCGAAGCTCTT
2H1		ATCATGCTCCAACAACAAGGAGATTCTGTGGGATGGGGAGCCTGTAGGCACAACATGTTCTGGCGAAGCTCTT
2P - Clone1		-----GGCACAAACATGTTCTGGCGAAGCTCTT
2P - Clone2		ATCATGCTCCAACAACAAGGAGATTCTGTGGGATGGGGAGCCTGTAGGCACAACATGTTCTGGCGAAGCTCTT
		2289

B

	716		762
GDDH13		DEENDNAAVTEQAQKPSSI SRQTKSCSNNKEILWDGEPVGTTCSGEA	
Gala		DEENDNAAVTEQAQKPSSI SRQTKSCSNNKEILWDGEPVGTTCSGEA	
2C2		DEENDNAAGHNMFWRSSLYACQSSWR*	
2C3		DEENDNAAGHNMFWRSSLYACQSSWR*	
2H1		DEENDNAAVNGTSPEA*	
2P - Clone1		DEENDNAAGHNMFWRSSLYACQSSWR*	
2P - Clone2		DEENDNAAVTEQAQKPSSISRQTKSCSNNKEILWDGEPVGTTCSGEA	

Figure 2.2-6: Analyse moléculaire des lignées mutantes. A. Alignement des résultats de séquençage de 3 lignées mutantes homozygotes et du séquençage du gène MD15G1231100 sur Gala sur la séquence de référence du génome GDDH13. En bleu la position des crRNAs. En vert les insertions, en rouge les délétions, en rose le polymorphisme entre la séquence de GDDH13 et de Gala. B. Alignement de la séquence protéique de chaque lignée mutante sur la séquence protéique de référence GDDH13 et de Gala. Est indiquée la position relative des crRNAs (bleu). En rose l'acide aminé correspondant au polymorphisme entre la séquence de Gala et de GDDH13. En jaune sont indiqués les codons STOP.

2.2.5. Phénotypage en serre des lignées mutantes

Une fois les lignées suffisamment développées nous avons entamé une procédure d'enracinement *in vitro*. Cette procédure occasionne un stress important des plantules, elle nécessite donc un matériel végétal en bonne santé sanitaire et en suffisamment grande quantité car une partie des plantes ne survivra pas.

Au total j'ai réussi à acclimater trois plantules de la lignée 2C3, cinq pour la lignée 2H1 et quinze pour la lignée 2P, je n'ai pas tenté d'acclimater de plantes de la lignée 2C2 car l'effectif de cette lignée était trop faible et l'acclimatation engendrerai des pertes trop importantes. Au cours des premiers mois de croissance en serre la quasi-totalité des plants se sont développés sans présenter de phénotype particulier (**Error! Reference source not found.A**). J'ai cependant observé qu'une seule des plantes de la lignée 2C3 a développé un phénotype différent, notamment au niveau de la forme de ses feuilles (**Error! Reference source not found.B**). On peut voir sur la photo que la feuille semble un peu gaufrée, et que la marge de la feuille est très altérée. C'est un phénotype que j'ai observé à partir de deux mois de croissance des plantes en serre et qui a perduré bien que nous ayons rabattus les plantes régulièrement. Ce phénotype n'est à priori pas lié à une maladie ou une infestation d'insectes, car aucune autre plante autour ne présentait un aspect similaire.

A



Gala « sauvage »

2H1



2C3



2P

B



Gala « sauvage »



2H1



2C3



2P

Figure 2.2-7 : photographies de plants des lignées acclimatées en serre. A. Photographie de 2 plantes par lignée sauvage ou éditée environ 1 mois après acclimatation. B. Photographie de 2 feuilles par lignée sauvage ou éditée environ 3 à 4 mois après acclimatation.

2.2.6. Caractérisation épigénétique des mutants

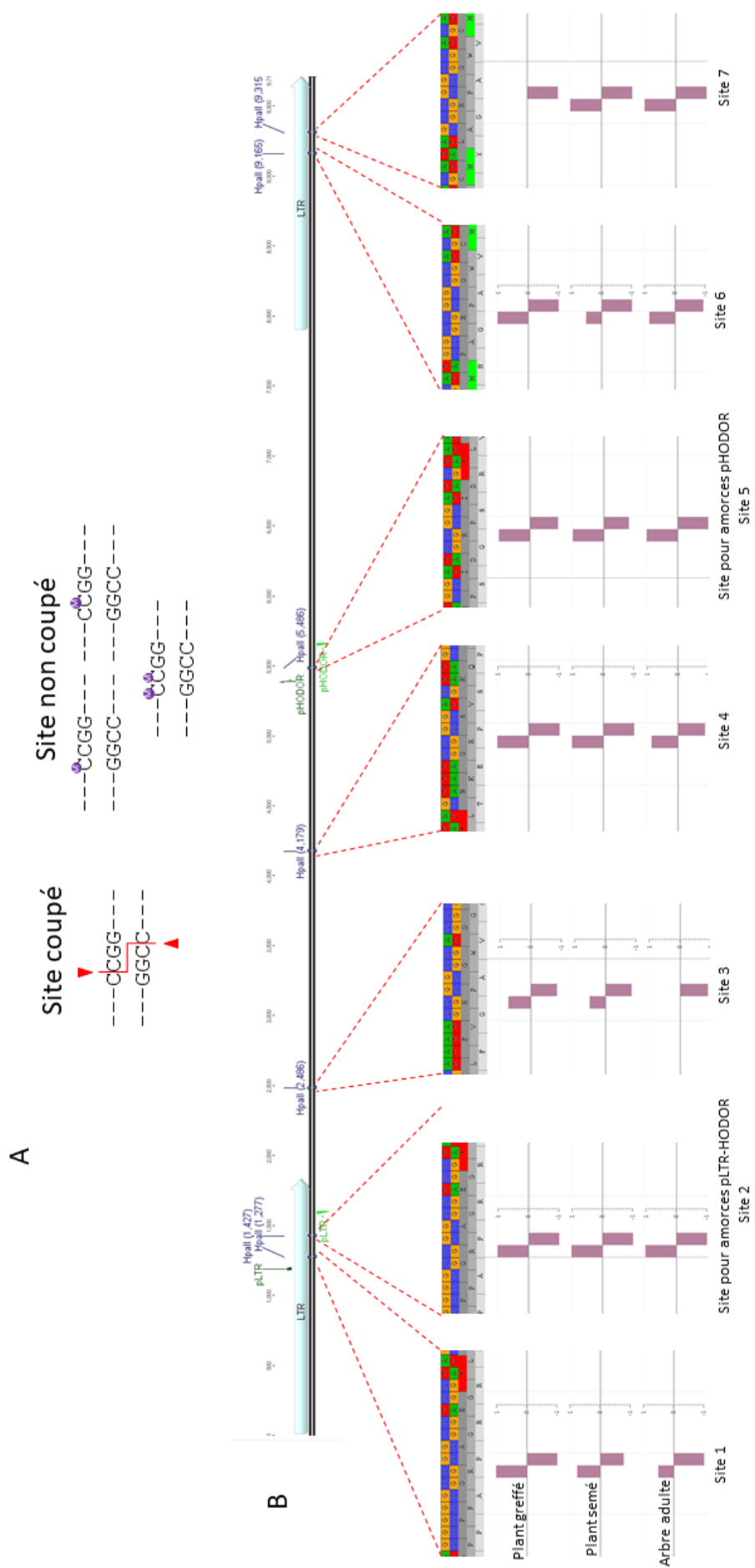
D'après les prédictions, les éditions du gène *Md15MET1* que nous avons réalisées doivent donner lieu à la formation de protéines MET1 tronquées et que l'on suppose non fonctionnelles. MET1 est impliqué dans la maintenance de la méthylation en contexte CG. D'après la littérature, les mutants *met1* chez *A.thaliana* (Kankel *et al.*, 2002) présentent une diminution du taux de méthylation en contexte CG. Nous pouvons donc supposer qu'une mutation du gène *Md15MET1* engendre également une diminution de la méthylation des cytosines en contexte CG chez le pommier.

J'ai cherché à mettre en évidence une diminution de la méthylation de deux façons. Tout d'abord par la mise en place d'un épigénotypage, une technique rapide qui utilise la PCR et la restriction enzymatique, et par la technique du séquençage bisulfite pour avoir une vision globale du statut de méthylation des lignées éditées.

a) Epigénotypage

Pour effectuer cet épigénotypage, comme Kankel *et al.* (2003), j'ai utilisé les propriétés de certaines enzymes et notamment de l'enzyme *HpaII* (McClelland *et al.*, 1994) qui peut cliver l'ADN uniquement si son site de coupure n'est pas méthylé (**Error! Reference source not found.A**). J'ai donc sélectionné une séquence d'ADN : l'élément transposable hautement répété dans le génome du pommier HODOR (High Copy Golden Delicious Repeat, LTR retrotransposon, Daccord *et al.* (2017)). HODOR portant des sites de restriction *HpaII*, qui sont supposés méthylés.

J'ai identifié sept sites *HpaII* sur l'élément transposable HODOR (**Error! Reference source not found.B**). A partir des données de méthylation précédemment obtenues sur différents échantillons (profil de méthylation d'échantillons de l'article Perrin *et al.* 2020, **Error! Reference source not found.B**). J'ai ensuite éliminé les sites 1, 3, 6 et 7 car la méthylation n'était pas maximale dans au moins un des échantillons. J'ai donc choisi les sites n°5 et n°2 autour desquels j'ai dessiné des amorces (respectivement pHODORs et pLTR-HODOR, annexe 6). Le site n°4 aurait également pu être utilisé. Lors des PCR de mises au point l'amplification s'est révélée plus facile pour les amorces pHODOR que pour celle pLTR-HODORs, j'ai donc réalisé la suite des tests avec les amorces pHODOR uniquement (site n°5 plutôt que 2).



Dans la **Error! Reference source not found.** sont présentés les résultats d'un des essais de digestion suivi d'une PCR réalisés. J'ai utilisé ici des extractions d'ADN : du témoin 'Gala' non transformé, de la lignée 2C3 (en prenant une plante avec (2C3#2) et sans phénotype (2C3#1) foliaire, confère partie 3.1.5), de la lignée 2H1 et de la lignée 2P.

Concernant le témoin de non-digestion (EF1 α), l'amplification est présente dans chaque échantillon et les variations d'intensité de bande sont *a priori* imputables à une différence de quantité de matrice utilisée initialement, bien que les concentrations initiales aient été ajustées par dosage au Qubit (fluoromètre de Invitrogene, Thermofisher). Pour les amplifications sur le transposon HODOR, à la suite de digestion enzymatique par *HpaII*, j'observe soit (i) pas de diminution comme pour le témoin Gala#2 soit (ii) une diminution de l'intensité, comme pour les témoins Gala#1 et 3 par rapport aux échantillons non digérés. Les échantillons ne présentant pas de diminution d'intensité pour l'amplification avec les amorces pHODORs à la suite de la digestion par *HpaII* sont éliminés pour la suite de l'analyse. Soit la digestion a mal fonctionné, soit la quantité d'ADN initial pour la PCR était trop importante.

Afin d'analyser mes résultats, j'ai mesuré l'intensité des bandes sur gel d'agarose à l'aide du logiciel ImageJ®. J'ai ensuite normalisé les intensités pour les amplifications sur le transposon HODOR par celles obtenues pour les amplifications sur le gène EF1 α (confère le matériel et méthode correspondant pour plus de détails), avant de réaliser un ratio des échantillons non-digérés sur digérés (**Error! Reference source not found.C**). Plus le ratio est élevé plus la différence d'intensité entre échantillons digérés et non-digérés est importante. Comme l'édition du gène *MET1* est supposée réduire le taux de méthylation, le site *HpaII* du transposon HODOR est supposé être plus accessible par l'enzyme dans les plantes éditées. L'amplification PCR pour cette zone est donc sensée être moins importante que dans le témoin non édité. Le ratio des échantillons non-digérés sur digérés pour les plantes éditées est donc supposé être de valeur supérieure à celui du témoin non transformé. Nos résultats semblent indiquer que tel est le cas pour les mutants 2C3#1, 2C3# et 2H1. Le mutant 2P ne présente a priori pas de différences avec le témoin non transformé 'Gala'.

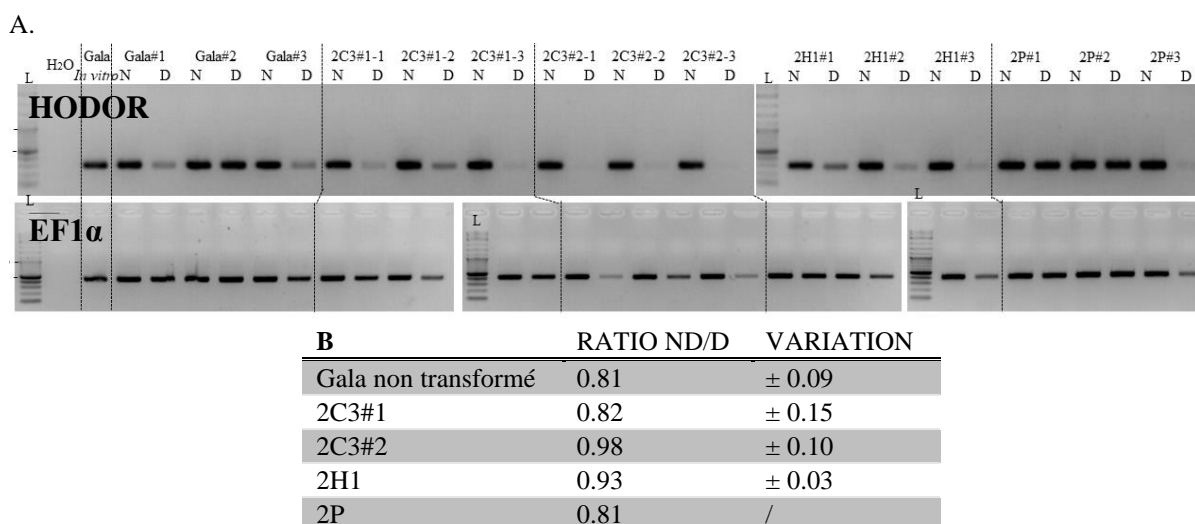


Figure 2.2-9 : Epigénotypage des lignées mutantes pour le gène MD15G1231100. A. PCRs sur le transposon HODOR et le gène de ménage EF1 α , sur de l'ADN non digéré (N) et digéré (D) par *HpaII*, des lignées mutantes et du témoin 'Gala' non transformé. En B. ratio des intensités des bandes HODOR digéré sur non-digéré. Les intensités HODOR ont préalablement été normalisées en fonction de celles obtenues pour les bandes Ef1 α . Un ratio supérieur à celui de 'gala' non transformé indique que l'ADN a subi une digestion plus importante dans la lignée mutante considérée par rapport au témoin 'Gala' non transformé.

b) Séquençage bisulfite du génome complet

Les digestions enzymatiques réalisées précédemment à l'aide d'une enzyme sensible à la méthylation nous ont donné des indices sur une potentielle diminution de la méthylation sur le transposon HODOR au sein des lignées mutantes 2C3 et 2H1 pour le gène *Md15MET1*. Nous avons donc décidé de réaliser un séquençage bisulfite pour analyser le niveau de méthylation global des lignées. Pour cette partie j'ai choisi de réaliser un séquençage bisulfite sur les lignées mutantes présentant un phénotype stable sur l'ensemble de la lignée, j'ai donc écarté la lignée 2C3 (confère partie 3.1.5). Nous avons également utilisé le mutant hétérozygote 2P, qui ne présentait pas de diminution de la méthylation dans le test précédent. Cependant, il est possible chez cette plante que les deux allèles (sauvage et édité) s'expriment en même temps, ce qui pourrait engendrer une diminution partielle de la méthylation par rapport aux mutants homozygotes. Cette diminution partielle de la méthylation est non indentifiable par la méthode semi-quantitative via *HpaII*.

2.3. Discussion

2.3.1. Efficacité de transformation

Je n'ai obtenu des pommiers édités pour la fonction MET1 que pour le gène *Md15MET1* (MD15G1231100) sur le génotype 'Gala'. Le taux de transformation total obtenu sur 'Gala' pour le gène *Md15MET1* est de 1.4% (transformations n°2 et 8 réunis). Ce taux aurait pu être plus élevé car nous avons stoppé la culture des cals une fois que nous avons suffisamment de lignées transformées. Le taux de transformation que nous avons obtenu est concordant avec ce qui a été réalisé par Charrier et al. (2019) dont les taux de transformation étaient de 0.75% à 14.25% pour le génotype 'Gala'.

a) Tentative de transformation du génotype GDDH13

Un de mes objectifs pour le projet BEEP était de transformer le génotype GDDH13. Maîtriser la transformation du génotype séquencé présente des avantages. Concernant le développement de la technique CRISPR-Cas9, si l'on dessine des crRNAs sur le génotype GDDH13 pour réaliser une transformation sur un autre génotype, il est nécessaire de vérifier la séquence du génotype à transformer notamment que les crRNAs ne présenteront pas de mésappariements. Cette vérification ne serait nécessaire si l'on transformait directement GDDH13.

Malgré les quatre transformations réalisées (N° 4, 5, 9 et 10, **Error! Reference source not found.**), je n'ai pas réussi à obtenir de lignées transformées pour le génotype GDDH13.

Plusieurs facteurs peuvent expliquer ces échecs. Tout d'abord l'état physiologique des plantes semble avoir un impact. Pour réaliser une transformation il est primordial que les feuilles prélevées soient saines et dans un état physiologique permettant leur réceptivité au protocole de transformation. Cet état est généralement atteint au bout de quatre semaines de croissance *in vitro*. Au-delà les vitro-plants s'épuisent sur un milieu manquant de certaines ressources, il peut en résulter un déséquilibre de la flore interne et la résurgence de bactéries endogènes au pied des vitro-plants. Or pour les transformations N° 9 et 10, les feuilles ont été prélevées au bout de cinq semaines de croissance. Les feuilles avaient un bel aspect mais elles présentaient peut-être déjà un déséquilibre au niveau de leur flore interne car nous avons constaté par la suite un développement bactérien important que nous n'avons pas observé lors des transformations N° 4 et 5.

Ensuite, les vitro-plants de GDDH13 ont été multipliés sur un milieu POM qui est optimisé pour la culture *in vitro* du génotype ‘Gala’. J’ai pu observer que les vitro-plants de GDDH13 avaient plus de mal à se développer sur ce milieu que ceux de ‘Gala’. Modifier la composition du milieu de culture et l’adapter à GDDH13 aurait peut-être permis d’améliorer les rendements de transformation pour ce génotype.

Lors des transformations N° 4 et 5, j’ai obtenu des bourgeons qui n’ont pas survécus. Il est possible que ces bourgeons n’aient pas été transformés et se soient développés sur un cal, lui-même transformé. Ce cal réalisant la détoxification de la kanamycine pour les bourgeons non transformé qu’il porte, leur permettant ainsi de se développer. Ce phénomène est fréquent lors de la transformation du pommier. Ces bourgeons « échappés », une fois isolés et placés en contact direct avec le milieu sélectif subiraient la sélection par la kanamycine et seraient détruits.

2.3.2. Un taux important de profil d’édition homozygote

Au début du projet BEEP, nous nous attendions à obtenir des profils de mutations très variés, comprenant des chimères d’édition avec des mélanges de profils d’éditions hétérozygotes, bi-alléliques mais très peu de profils homozygotes comme cela a été montré par Charrier et al. (2019),. Les chimères d’édition sont définies comme des plantes régénérées à partir de quelques cellules ayant toutes intégré la cassette CRISPR-Cas9 mais dans lesquelles la Cas9 n’a pas agi au même moment et où les réparations ont donc été différentes. Ces plantes sont donc totalement transformées mais édités différemment d’une cellule à l’autre. Les profils d’édition pouvaient être aussi nombreux que le nombre de cellules initialement impliquées dans la régénération. La complexité augmentait encore si l’une de ces cellules initiales n’était pas encore éditée. La lignée cellulaire correspondante donne alors lieu à de multiples profils d’édition différents. Les mutants bi-alléliques sont des mutants pour lequel des mutations différentes sont observées sur les deux allèles du gène.

Or sur onze lignées éditées isolées, j’ai identifié deux (2C3 et 2H1) lignées au profil d’édition homozygote en T0. C’est un taux élevé (18%) comparé aux précédents travaux sur pommier et poirier (~2%, Charrier et al 2019) ou sur d’autres ligneux tels que la vigne (0%, Nakajima et al. (2017); Wang et al. (2018)) et l’oranger (5,3%, Peng et al. (2017)).

Ces résultats nous permettent de supposer que les événements d’éditions chez ces lignées ont eu lieu tôt au cours du processus de transformation et que les cellules les portant étaient donc nombreuses dans les cals organogènes, limitant ainsi le nombre de plantes présentant des chimères d’édition et augmentant le taux d’éditions homozygotes. Le gène *MET1* a pour fonction la maintenance de la méthylation (Finnegan *et al.*, 1996, 1998) en contexte CG. Ce gène est exprimé durant la multiplication cellulaire afin d’assurer la méthylation des brins de chromatines nouvellement formés durant le cycle cellulaire (Desvoyes *et al.*, 2014; Liu *et al.*, 2018). Le gène *MET1* est donc exprimé précocement pendant l’organogénèse (*e.g.* : comme dans un cal, Li et al. (2014); Yamamoto et al. (2005)). Son accessibilité par la machinerie transcriptionnelle pendant l’organogénèse laisse sous-entendre qu’il est sans doute dans le même temps facilement accessible par la machinerie CRISPR-Cas9. Ceci peut expliquer une édition précoce de ce gène au cours du processus de transformation.

2.3.3. La lignée 2P un hétérozygote pour toujours ?

Lors du séquençage des clones bactériens réalisés pour le génotypage de la lignée 2P, nous avons identifié qu'un des allèles du gène *Md15MET1* était toujours sauvage. Nous avons cependant choisi de conserver un tel mutant dans notre expérimentation car le système CRISPR-Cas9 est toujours présent dans les cellules de cette lignée et peut continuer à éditer cet allèle non muté. Il est probable que la lignée 2P finisse par devenir totalement éditée au gène cible, mais vraisemblablement sous la forme d'une chimère d'édition.

2.3.4. Les lignées 2C2 et 2C3 sont peut-être une seule et même lignée

Les lignées 2C2 et 2C3 présentent exactement le même évènement d'édition et sont issues de bourgeons isolés de la même feuille. Il est possible que ces deux lignées proviennent également du même cal organogène. Si elles proviennent du même cal, alors ces lignées peuvent être le résultat d'un évènement d'édition unique ayant eu lieu très tôt au cours du processus de régénération, comme expliqué précédemment (partie 3.2.2.). Si en revanche ces deux lignées sont issues de calcs différents alors même si le profil d'édition est identique, ces deux lignées sont à considérer comme deux évènements de transformations différents, car la cassette d'édition a probablement été insérée à des loci différents dans le génome. Le lieu d'insertion de la cassette pourra être ultérieurement recherché via différentes méthodes (Haraguchi & Nakagawara, 2009; Moisan, 2010; Thirulogachandar *et al.*, 2011). Ces méthodes permettent, à partir du transgène, de séquencer les bordures et ainsi de définir la localisation de l'insertion.

2.3.5. Séquençage bisulfite et test de digestion

Afin d'analyser si les mutants de *MdMET1* obtenus sur pommier présentent une diminution de la méthylation de l'ADN comme observé pour *A.thaliana* (Bartee & Bender, 2001), nous avons réalisé dans un premier temps un test de digestion enzymatique avec une enzyme sensible à la méthylation, sur un élément transposable hautement répété dans le génome du pommier. Les résultats suggèrent une faible diminution de la méthylation en contexte CG sur les transposons HODOR pour nos mutants.

Face à l'imprécision de cette technique et aux difficultés rencontrées, nous avons choisi d'effectuer un séquençage bisulfite sur les mutants 2H1 et 2P, afin d'obtenir le profil de méthylation sur l'ensemble du génome. Il sera intéressant de vérifier si cette mutation a été compensée par une autre voie de méthylation de l'ADN, si le profil de méthylation est inchangé, ou si certaines zones du génome sont plus appauvries en méthylation que d'autres.

Dans la littérature, de nombreuses données de séquençage bisulfite ont été obtenues pour des mutants du gène *MET1*, chez *A.thaliana*, *Oryza sativa* ou *Solanum lycopersicum*. Le Tableau 2.3-1 présente une partie de ces données. Dans tous les mutants il est possible d'observer une diminution de la méthylation en contexte CG. La méthylation en contexte CHG et CHH varie quant à elle différemment. Au sein d'une même espèce, certaines études rapportent une diminution de la méthylation dans ces deux contextes, quand d'autres études rapportent une augmentation de la méthylation. Il a également été constaté des modifications phénotypiques chez ces plantes, notamment au niveau de la floraison, qui est retardée, mais également au niveau des feuilles.

De ce fait pour les mutants du le gène *MET1* chez le pommier, nous nous attendons à observer une diminution importante de la méthylation en contexte CG. Nous devrions également pouvoir observer des modifications des niveaux de méthylation dans les contextes CHG et CHH. Pour ces deux contextes, les modifications des niveaux de méthylation ne semblent pas prévisibles et pourraient dépendre du contexte génétique. L'observation d'un phénotype de retard de floraison ne sera pas possible dans le temps de l'étude, d'autant plus si le temps de floraison est retardé. Il sera cependant possible d'observer le développement d'un phénotype foliaire le cas échéant.

Tableau 2.3-1 : Caractéristiques des mutants *met1* de trois espèces différentes

Espèce	CG	CHG	CHH	Phénotype	Références
Arabidopsis thaliana	↘↘	↘	↘		Kankel et al., (2002)
	↘↘	↘ (maintenance) ↗ (de novo)	↘		Aufsatz et al (2004)
	↘	↘	↗	Floraison retardé Plus de feuilles	Mathieu et al., (2007)
	↘↘	↗ (euchromatine) ↘ ("DMR")	↘		Lister et al., (2008)
			↘ ("DMR")		
Oryza sativa	↘	↘	↘	Semences anormales	Yamauchi et al., 2014
	↘	↘	↗ (DMR)	Développement ralenti / anormal	Hu et al., (2014)
	↘ (DMR)	↘ (DMR)			
Solanum lycopersicum	↘	↘	↗	Feuille enroulés	
				Moins d'inflorescence	Yang et al., 2019
				Moins de fleur / inflo	

Les échantillons ont été envoyés pour un séquençage de méthylome à la plateforme de Génome Québec en novembre 2019. L'analyse préliminaire sera réalisée et présentée lors de la soutenance en mai 2020 et donnera lieu à publication ultérieurement. La première étape de l'analyse consistera à extraire le niveau moyen de méthylation de chaque contexte pour chaque échantillon afin d'évaluer l'évolution du taux de méthylation entre le 'Gala' sauvage et les lignées mutantes 2H1 et 2P. Un alignement des séquences bisulfites sera ensuite réalisé afin de calculer les régions différenciellement méthylées. Cette analyse permettra d'identifier des régions où la méthylation serait particulièrement affectée par la mutation du gène *MET1*.

2.3.6. Phénotypes attendus pour des plantes dont les voies de maintien de la méthylation sont défectueuses

Il a été montré chez *A.thaliana* que les plantes présentant un défaut d'une voie de méthylation de l'ADN sont viables et ne présente pas d'infertilité, mais accumulent des modifications phénotypiques au cours des générations (Finnegan *et al.*, 1996; Kakutani *et al.*, 1996; Ronemus *et al.*, 1996b; Richards, 1997). J'ai observé un phénotype particulier seulement pour l'une des plantes de la lignée 2C3. J'ai essayé d'augmenter la probabilité d'apparition de phénotypes en rabattant régulièrement les plantes pour favoriser leur croissance et en réalisant de la multiplication asexuée par micro-greffage. Malheureusement, ces méthodes ne m'ont pas permis d'observer d'autres phénotypes particuliers. Il faudrait peut-être réaliser de la multiplication sexuée sur plusieurs générations et ainsi se mettre dans les mêmes conditions que ce qui a précédemment été fait chez

A.thaliana pour obtenir une accumulation de phénotypes, mais cela reste difficilement réalisable sur pommier car chaque génération prend plusieurs années du fait d'une période juvénile particulièrement longue (Visser, 1964).

Il est cependant envisageable de réaliser deux ou trois années de greffage consécutives sur ces lignées, afin d'étudier l'évolution du taux de méthylation lors d'une multiplication asexuée, et son éventuel impact phénotypique.

3. Edition de l'épigénome du pommier – modification de la méthylation d'un gène pour modifier son expression

3.1. Choix méthodologiques en vue de l'obtention d'une construction permettant la méthylation d'une séquence cible

Nous allons ici détailler différents protocoles que nous avons mis en place pour obtenir les constructions nécessaires à la modification de la méthylation du gène *MdTFL1*.

3.1.1 Choix des crRNA ciblant le promoteur de *MdTFL1*

Après observation de la méthylation du gène *MdTFL1* sur les données du séquençage bisulfite obtenues précédemment (cf : Chapitre I) et afin de choisir les meilleures zones à cibler, une analyse des crRNAs possibles est réalisée à l'aide du logiciel CRISPOR-Tefor présenté précédemment (cf : **Error! Reference source not found.**). Chaque crRNA est ensuite inséré *in silico* dans un gRNA et synthétisé comme vu précédemment (partie 2.1 matériel et méthodes). J'ai également prévu des sites de restriction de part et d'autre des gRNAs afin de les associer dans la construction finale via la technique du Golden Gate.

3.1.2. Choix de la méthyltransférase et de son promoteur

Afin de méthyler le promoteur du gène *MdTFL1* nous avons sélectionné le gène DOMAINS REARRANGED METHYLASE 2 (DRM2) provenant d'*A.thaliana* (**Error! Reference source not found.A**, AT5G14620, TAIR (2020), <https://www.arabidopsis.org/servlets/TairObject?id=138028&type=locus>). Cette enzyme a été choisie pour sa capacité à réaliser de la méthylation *de novo* (cf : Introduction partie 4.3.3 a)) et à transférer de la méthylation sur les trois contextes de cytosine. Trois portions du gène sont utilisées pour réaliser trois constructions: la séquence codante complète (DRM2T, **Error! Reference source not found.A**), la séquence codant pour la fonction méthyltransférase seule (DRM2S, **Error! Reference source not found.A**), et la séquence codant pour la fonction méthyltransférase modifiée pour l'inactiver (DRM2KO, **Error! Reference source not found.B**). Cette modification a été réalisée d'après l'analyse de la fonction d'AtDRM2 réalisée en 2014 par Zhong et al. Cette dernière séquence peut être utilisée pour réaliser un témoin de transformation. Dans Charrier et al. (2019), il a été montré que l'obtention de plants édités pour le gène *MdTFL1*, via l'activité d'une Cas9 sous contrôle d'un promoteur 35S du CAMV, permet l'observation de floraisons à des stades très précoces après régénération. Mais la floraison inhibe la croissance de ces plants et il est alors très difficile, voire impossible de les maintenir. Pour pallier ce problème nous avons choisi de placer la méthyltransférase (DRM2) sous contrôle d'un promoteur inductible à la chaleur du soja (HSP, **Error! Reference source not found.B**, Schöffl, Raschke, and Nagao (1984)) et déjà validé par l'équipe ResPom comme fonctionnel chez le pommier.

L'utilisation de ce promoteur permettra ainsi d'obtenir des plants ayant intégré la cassette de transformation et de les faire se développer suffisamment pour les acclimater en serre. Une fois

développés, il est possible de les soumettre à un choc thermique pour activer l'expression de la cassette et ainsi obtenir une floraison.

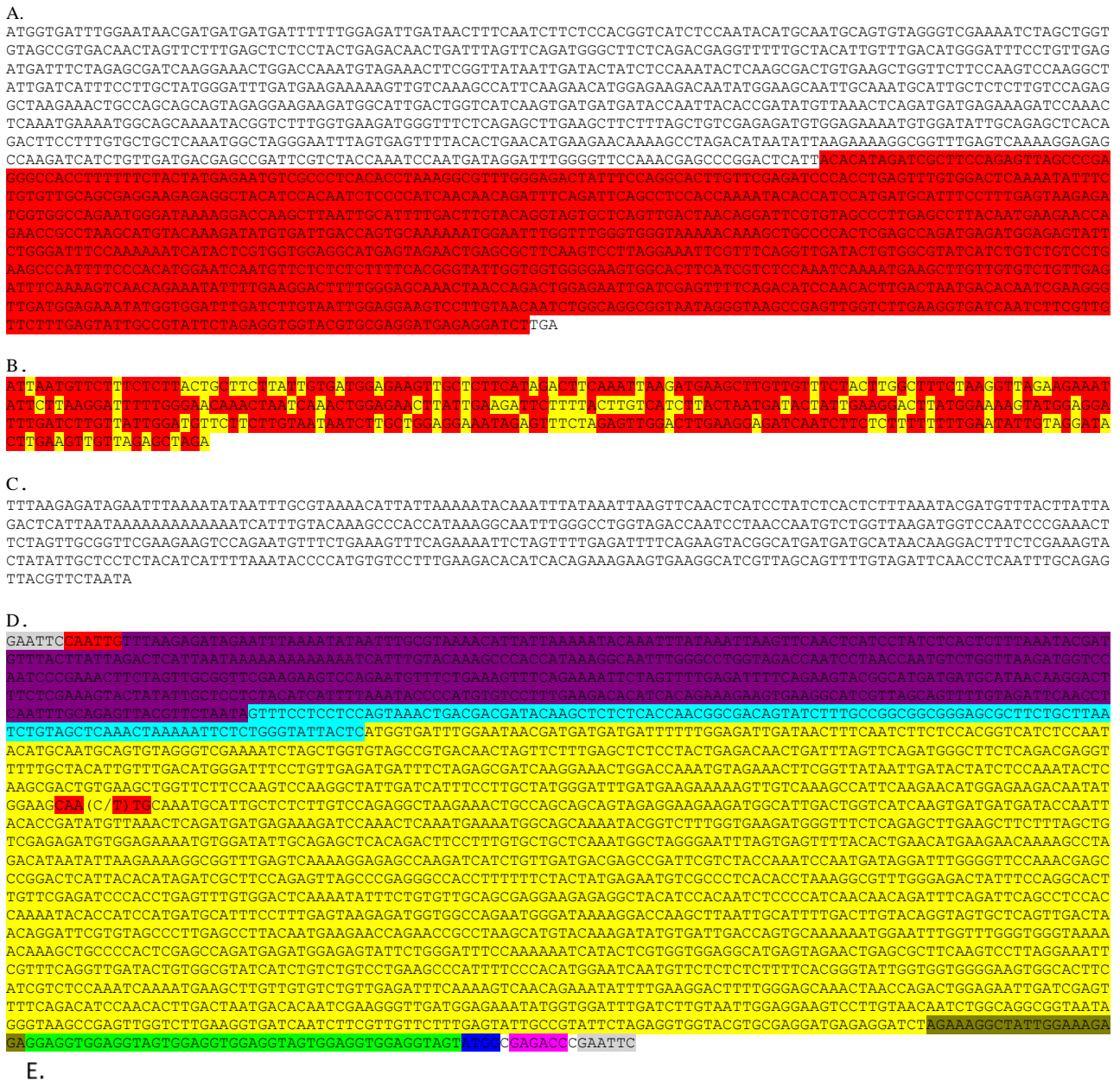


Figure 3.1-1 : Description des différentes séquences d'ADN utilisées pour la partie d'édition du promoteur du gène Tfl1. A. séquence du gène codant pour la fonction DMR2 chez *A. thaliana*. En rouge la partie annotée comme portant la fonction méthyltransférase. B. Séquence nucléotidique de DMR2 annotée comme portant la méthyltransférase avec les modifications nucléotidiques réalisées (en jaune) pour inactiver la fonction méthyltransférase (DRM2KO), permettant la production de plantes transformées témoins validant l'activité méthylase de DRM2 dans les lignées transgéniques pour les autres constructions. C. séquence d'un promoteur inductible à la chaleur du soja (Schöffl et al, The EMBO J, 1984). D. construction HSP::DRM2. En rouge, rose et gris les sites de restriction *MfeI*, *BsaI* et *EcoRI*, respectivement. En violet la séquence HSP, en bleu clair la région 5'UTR du gène *AtDRM2*, en jaune la séquence codante du gène *AtDRM2* (ici la séquence complète), en jaune foncé une séquence d'adressage

au noyau, en vert trois séquences de liaison GS, en bleu foncé quatre nucléotides correspondant au début de la séquence codante de la Cas9 pour réaliser un ligation Golden Gate. E. représentation schématique de la séquence D.

3.1.3 Choix et obtention de la Cas9 inactivée

Dans cette partie, les clonages réalisés étaient compliqués de par la taille des insertions à réaliser. Plusieurs protocoles étaient possibles pour obtenir un plasmide contenant une Cas9 inactivité (pDE-deadCas9-kr, **Error! Reference source not found.**). Les protocoles présentés ici sont disponibles en annexe, ils ont été adaptés à partir de différents protocoles de clonage, de discussions sur les forums (Research Gate, Open Wet Ware ou NCBI (Laible & Boonrod, 2009)) et de protocoles et recommandations d'utilisations fournis avec les enzymes et réactifs utilisés (Neb et Promega).

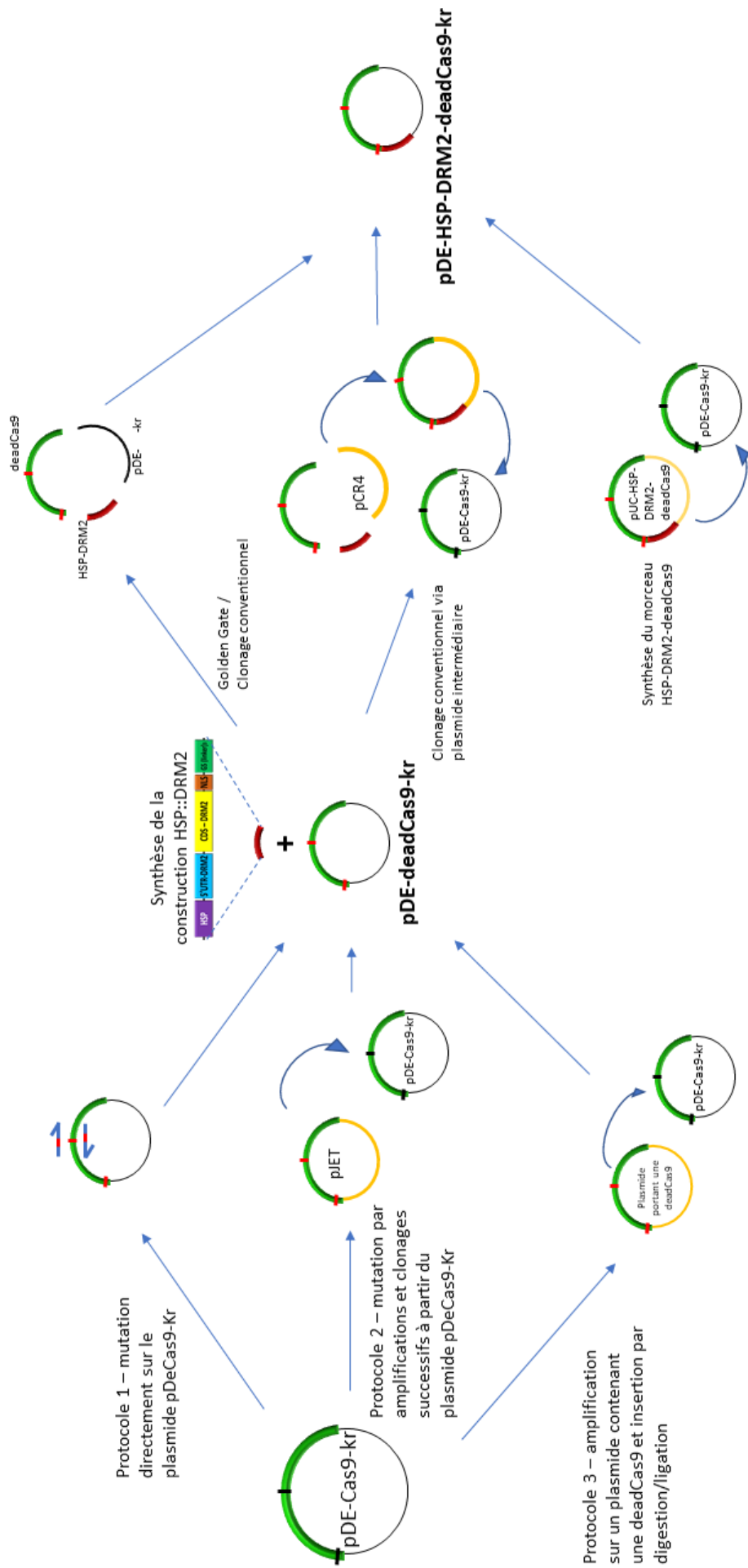


Figure 3.1-2: Schéma des protocoles possibles pour l'obtention du plasmide contenant la construction pDE-deadCas9-HSP-DRM2-kr. La mutation d' un site est schématisée par la couleur rouge

a) Premier protocole : mutation directement sur le plasmide pDeCas9-Kr (Error! Reference source not found. et Annexe 9)

Le plasmide pDE-Cas9-kr comporte une Cas9 active. L'objectif de ce protocole est d'obtenir une Cas9 inactive à partir de cette Cas9 active, en réalisant les mutations permettant les modifications d'acides aminés en position 10 et 840 de la séquence protéique. L'objectif est d'amplifier le plasmide entier à l'aide d'amorces dégénérées permettant d'ajouter les mutations aux sites choisis. On dessine tout d'abord des couples d'amorces dont les amorces en sens et en antisens sont complémentaires et on intègre un changement nucléotidique permettant de modifier les acides aminés ciblés en alanine (amorces N° 23 à 26, annexe 6).

Les changements nucléotidiques apportés permettent de modifier une adénine en cytosine dans le codon correspondant à l'acide aspartique situé en dixième position de la séquence protéique. Une cytosine et une adénine sont respectivement modifiées en guanine et cytosine dans le codon correspondant à l'histidine située en huit-cents quarantième position de la séquence protéique.

Puisque deux sites doivent être mutés, le protocole décrit (Annexe 9) est réalisé deux fois avec soit le premier couple d'amorces, soit le second. Pour l'amplification complète du plasmide, qui a une taille de 16 kb, la polymérase haute-fidélité « Q5 » (NEB) est utilisée et permet une amplification rapide (20 à 30 secondes par kb) de fragment allant jusqu'à 20kb avec un taux d'erreur plus faible que les polymérases habituelles (280 fois plus faible d'après les spécifications). L'amplicon produit présente des extrémités aux bordures franches, permettant au plasmide de se refermer sans étape supplémentaire.

Une étape de digestion est ensuite effectuée pour éliminer les plasmides natifs via l'enzyme *DpnI*, qui va cliver les séquences d'ADN méthylées. Le plasmide modifié est néosynthétisé pendant la PCR. Il n'est pas méthylé et n'est donc *a priori* pas dégradé. Le digestat obtenu est donc supposé contenir des morceaux du plasmide ayant servi de matrice et le plasmide portant la mutation, entier. Ce digestat est alors utilisé pour réaliser une transformation de bactéries et multiplier le plasmide. Puis le protocole est réitéré avec le second couple d'amorces pour muter le second site ciblé.

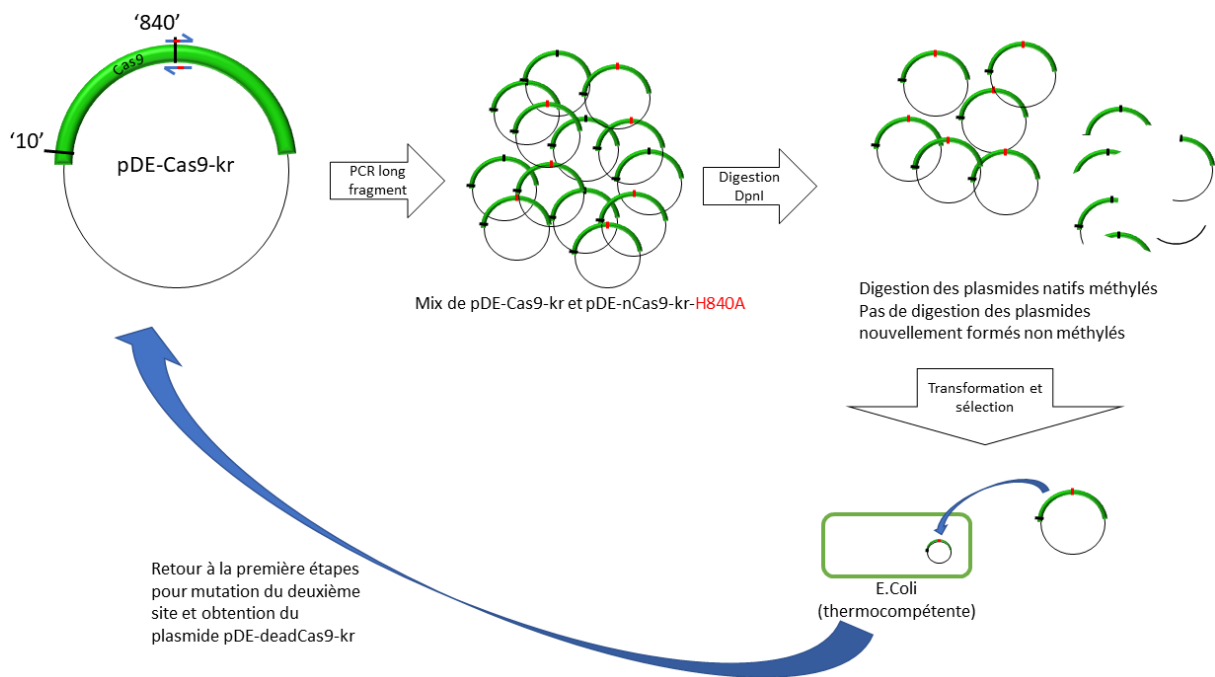


Figure 3.1-3: Schéma d'obtention des mutations nécessaires pour obtenir une deadCas9 à partir du plasmide pDE-Cas9-kr. La Cas9 est représentée en vert. Les sites de mutation sont indiqués par '10' et '840'. La mutation d'un site est schématisée par la couleur rouge. La description du protocole utilisé est disponible en Annexe 9.

b) Second protocole : mutation par amplifications et clonages successifs à partir du plasmide pDeCas9-Kr (*Error! Reference source not found.*, Annexe 10)

Le but de ce protocole est d'obtenir la Cas9 inactivée par amplifications et clonages successifs, toujours à partir de la Cas9 active sur le plasmide pDeCas9-Kr. La PCR permet d'amplifier deux fragments avec des amorces dégénérées (fragment 1 : amorce N°25 + 24, fragment 2 : amorces N°23 + 29, annexe 6), permettant l'ajout des mutations. Une seconde PCR permet l'ajout de sites *BsaI* en 3' du fragment 1 et en 5' du fragment 2 (les autres extrémités comportant un site *AscI* présent sur le plasmide initial). Ces sites permettent à l'étape suivante de relier les deux fragments produits en utilisant la technique du Golden Gate. Après ligation le site *BsaI* n'est plus présent au niveau de la ligation (cf *Error! Reference source not found.*C). Cette construction est ensuite clonée dans un plasmide intermédiaire, le pJET1.2 (Thermofisher) puis amplifiée par transformation bactérienne (bactérie Top 10 thermo-compétente (Thermofisher) ou bactérie électro-compétente (Annexe 11)) pour optimiser la transformation. Ce plasmide pJET1.2 est ensuite digéré avec l'enzyme *AscI* pour libérer la Cas9 mutée. Le plasmide pDE-Cas9-kr d'origine est également digéré avec l'enzyme *AscI*, pour sortir la Cas9 non mutée. Le plasmide pDE-Cas9-kr digéré au site *AscI* est ensuite déphosphorylé pour empêcher qu'il ne se referme sur lui-même ou ne réintègre la Cas9 non mutée. Enfin les deux produits de digestion sont assemblés grâce à une étape de ligation. La digestion/ligation Golden Gate entre les deux fragments PCR (fragments 1 et 2) peut aussi être tentée en même temps que l'insertion dans le plasmide final. Pour cela les deux fragments sont digérés de manière concomitante avec les enzymes *AscI* et *BsaI*, avant d'être assemblés lors de l'étape de ligation qui permet aussi de les introduire dans le plasmide. Les étapes de ligation sont réalisées pendant 48h à 4°C et en utilisant différentes concentrations d'inserts et de plasmide, afin d'optimiser les chances d'obtenir une construction complète.

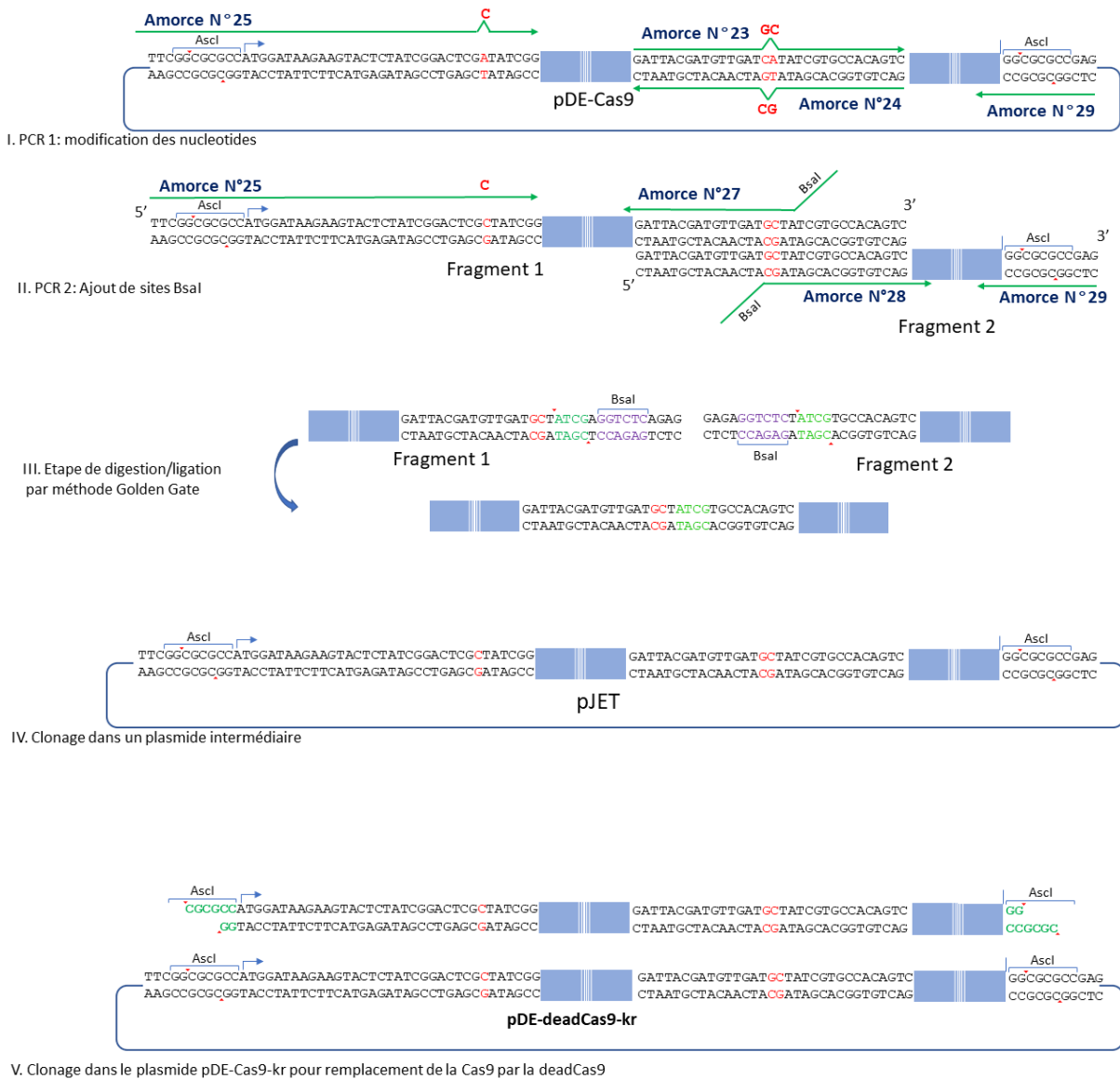


Figure 3.1-4: Schéma d'un second protocole permettant d'obtenir le plasmide pDEdeadCas9. Voir protocole détaillé en Annexe 10. Nucléotide en rouge: mutation désirée. Nucléotide en vert: motif de coupure de l'enzyme *BsaI* permettant une ligation spécifique des deux fragments. En violet site de reconnaissance de l'enzyme *BsaI*.

c) Troisième protocole : utilisation d'une deadCas9 d'un plasmide existant (*Error! Reference source not found.*)

A partir d'un plasmide contenant une Cas9 inactivée, la Cas9 inactivée est amplifiée avec les amorces N°25 et 29 (annexe 6) pour ajouter les sites *AscI* aux extrémités. Le plasmide pDE-Cas9-kr et l'amplicon sont digérés par l'enzyme *AscI*. Puis une ligation est réalisée avant transformation bactérienne.

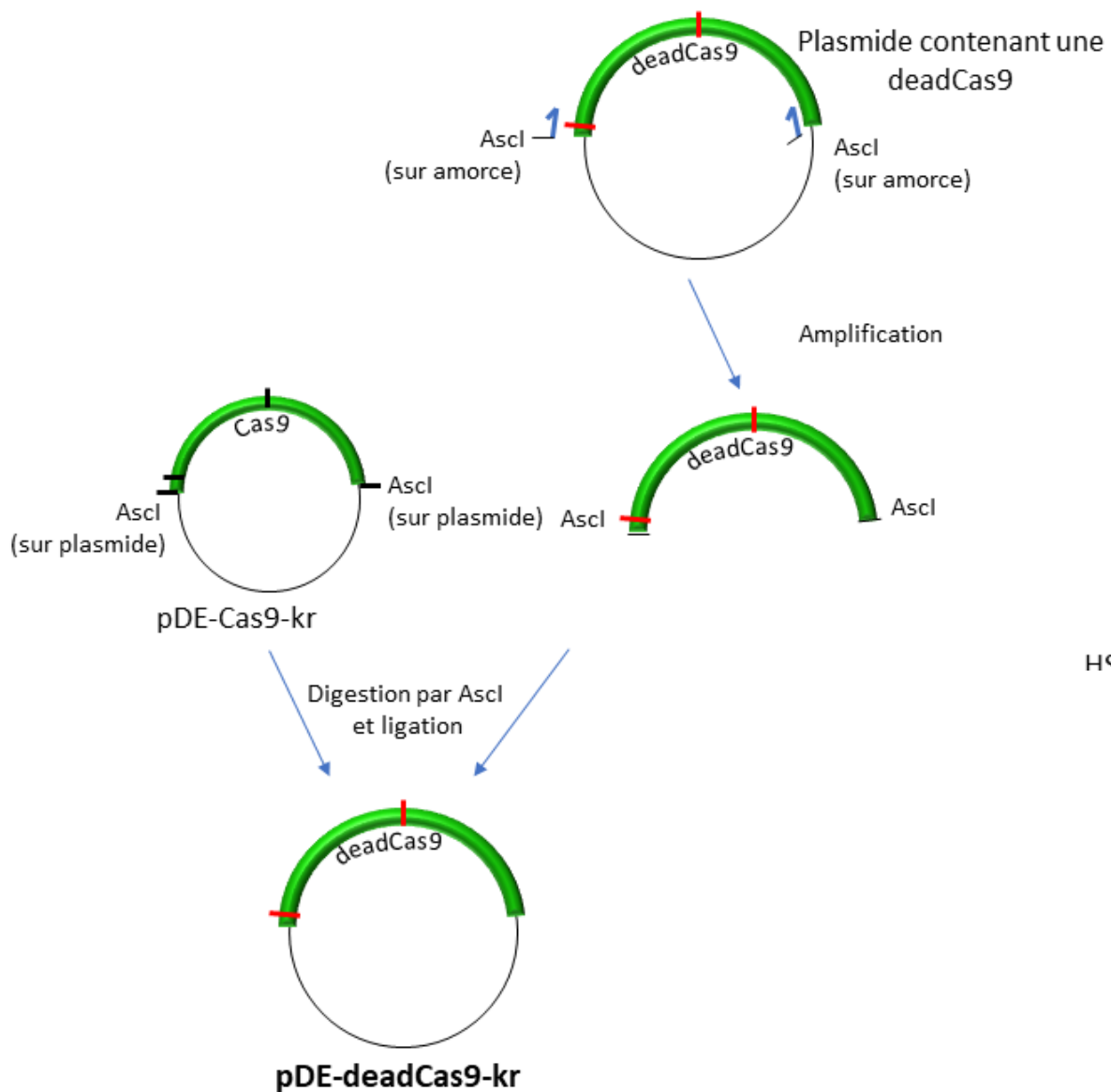


Figure 3.1-5 : Schéma du protocole 3

3.1.2. Obtention de la portion HSP::DRM2 (Error! Reference source not found.E)

A partir de la séquence du gène AT5G14620 codant pour la fonction DRM2 d'*A. thaliana* (Error! Reference source not found.A ; obtenue sur le site TAIR <https://www.arabidopsis.org/servlets/TairObject?id=138028&type=locus>), nous avons réalisé des constructions HSP::DMR2 (DRM2S, DRM2KO ou DRM2T, Error! Reference source not found.A et B) *in silico*. Ces constructions ont ensuite été synthétisées par la société Proteogenix (France). Pour créer ces portions de construction, je me suis inspiré des travaux de Liu et al. (2016) et Vojta et al. (2016). Chaque construction se compose (Error! Reference source not found.D et E) d'un promoteur inductible (HSP), de la région génique 5'UTR du gène *AtDRM2*, de la séquence codante complète de *AtDRM2* (HSP::DRM2T) de son site actif uniquement (HSP::DRM2S) ou de la séquence du site actif inactivée (HSP::DRM2KO, construction témoin), d'un signal d'adressage au noyau (en plus de celui présent dans la séquence de la Cas9 inactivée), ainsi que des séquences de liaison (GS linker) utilisées pour augmenter la mobilité entre les protéines associées (ici DRM2T, KO ou S et la deadCas9). Des sites de restriction ont aussi été placés en 5' et 3' de la séquence pour les étapes de clonage (Figure 3.1-1D et E).

Pour DRM2S et DRM2KO, un codon start a été rajouté en respectant le cadre de lecture, et pour DRM2T, le codon stop de la séquence codante du gène a été retiré.

3.1.3. Obtention d'un plasmide contenant la construction HSP::DRM2::deadCas9 à partir, ou non, de pDe-deadCas9-Kr

De la même façon que pour l'obtention d'un plasmide contenant la Cas9 inactivée, nous avons réalisé plusieurs protocoles en parallèle afin d'obtenir un plasmide contenant la construction HSP::DRM2::deadCas9 (pDE- HSP-DRM2-deadCas9-kr ; Error! Reference source not found.).

a) Protocole d'obtention par Golden Gate (Error! Reference source not found.)

Pour ce protocole, la séquence de la deadCas9 est amplifiée à partir du plasmide pDE-deadCas9-kr avec les amorces N°30 et 29 (annexe 6) pour ajouter le site *BsaI* en 5' de la séquence. La séquence deadCas9 amplifiée est ensuite digérée par les enzymes *AscI* et *BsaI* avant d'être purifiée. Les constructions HSP::DMR2 (T, S ou KO) sont digérées par les enzymes *MfeI* et *BsaI* avant d'être purifiées. Le plasmide pDE-Cas9-kr est digéré par les enzymes *MfeI* et *AscI*, les enzymes *XhoI* et *BspHI* sont également utilisées pour digérer la Cas9 en 6 fragments qui sont ensuite éliminés par purification. Le plasmide pDE ouvert est ensuite déphosphorylé.

Les étapes de ligation associant le reste du plasmide pDE, la deadCas9 et un fragment HSP::DRM2 sont réalisées pendant 48h à 4°C et en utilisant différentes concentrations d'inserts et de plasmide, afin d'optimiser les chances d'obtenir une construction complète.

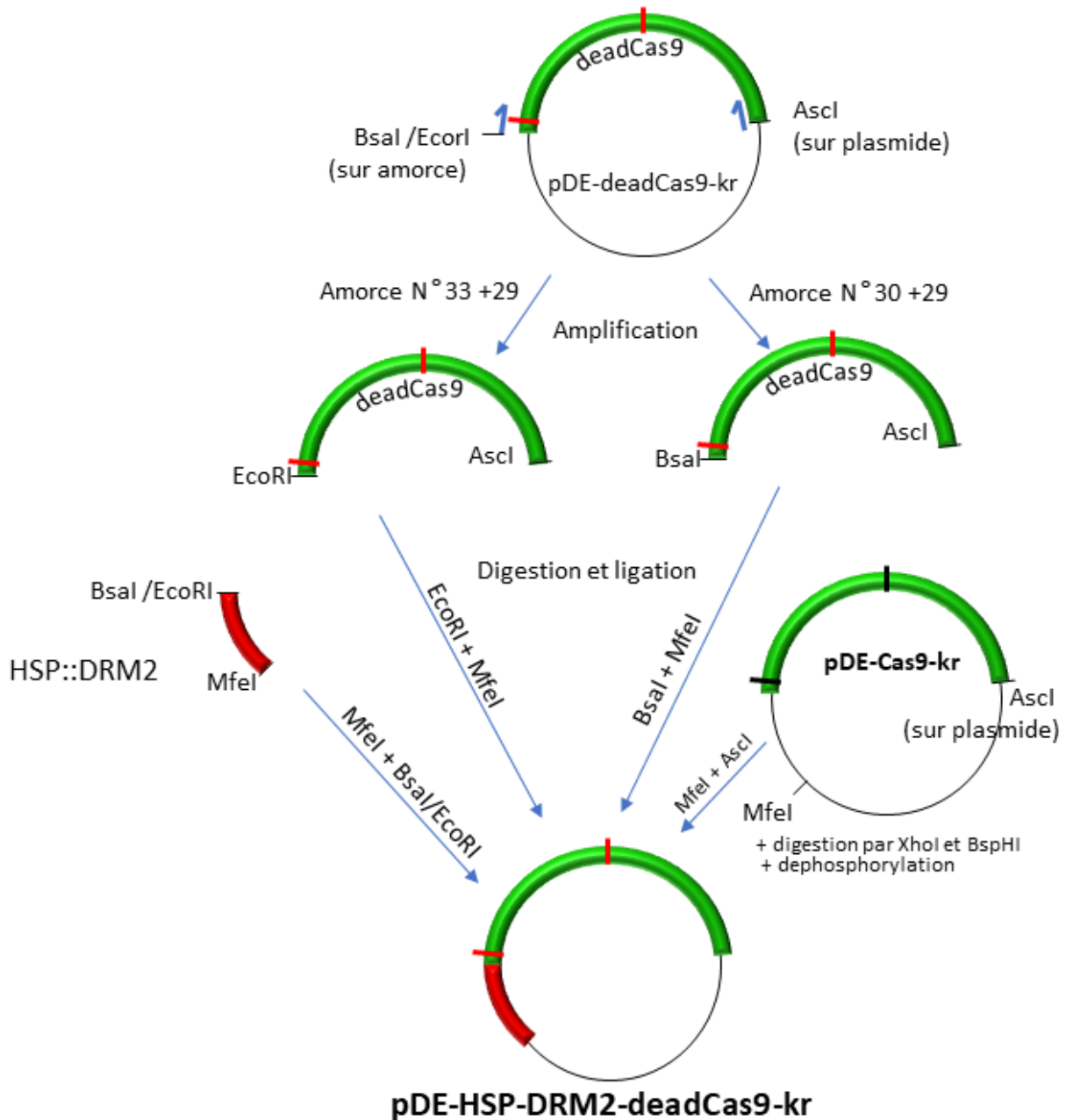


Figure 3.1-6 : Schéma du protocole 4

b) Protocole d'obtention par clonage conventionnel (*Error! Reference source not found.*)

Pour ce protocole la séquence de la deadCas9 est amplifiée à partir du plasmide pDE-deadCas9-kr avec les amorces N°33 et 29 (annexe 6) pour ajouter le site EcoRI en 5' de la séquence. La séquence deadCas9 amplifiée est ensuite digérée par les enzymes AscI et EcoRI avant d'être purifiée. Les constructions HSP::DMR2 (T, S ou KO) sont digérées par les enzymes MfeI et EcoRI avant d'être purifiées. Le plasmide pDE-Cas9-kr est digéré par les enzymes MfeI et AscI, les enzymes XhoI et BspHI sont également utilisées pour digérer la Cas9 en 6 fragments qui sont ensuite éliminés par purification. Le plasmide pDE ouvert est ensuite déphosphorylé. Les étapes de ligation associant le reste du plasmide pDE, la deadCas9 et un fragment HSP::DRM2 sont réalisées pendant 48h à 4°C et en utilisant différentes concentrations d'inserts et de plasmide, afin d'optimiser les chances d'obtenir une construction complète.

c) Protocole d'obtention à l'aide d'un plasmide intermédiaire (Error! Reference source not found.)

Pour ce protocole la séquence de la deadCas9 est amplifiée à partir du plasmide pDE-deadCas9-kr avec les amorces N°30 + 29 ou N°33 + 29 (annexe 6), pour ajouter respectivement le site BsaI ou EcoRI en 5' de la séquence. La séquence deadCas9 amplifiée est ensuite digérée par l'enzyme *BsaI* ou *EcoRI* avant d'être purifiée. Les constructions HSP::DRM2 (T, S ou KO) sont digérées par les enzymes *MfeI* et *BsaI* ou *EcoRI* avant d'être purifiées. Une ligation est effectuée pour obtenir la séquence HSP::DRM2::deadCas9, qui est ensuite clonée dans le plasmide pCR4 du kit de clonage « Zero Blunt » (Zero Blunt™ TOPO™ PCR Cloning Kit, Thermofisher). Le plasmide pCR4-HSP::DRM2::deadCas9 est ensuite digéré avec les enzymes *MfeI* et *AscI*. Le plasmide pDE-Cas9-kr est digéré par les enzymes *MfeI* et *AscI*, les enzymes *XhoI* et *BspHI* sont également utilisées pour digérer la Cas9 en 6 fragments qui sont ensuite éliminés par purification. Le plasmide pDE ouvert est ensuite déphosphorylé. Les étapes de ligation associant le reste du plasmide pDE et le fragment HSP::DRM2::deadCas9 sont réalisées pendant 48h à 4°C et en utilisant différentes concentrations d'inserts et de plasmide, afin d'optimiser les chances d'obtenir une construction complète.

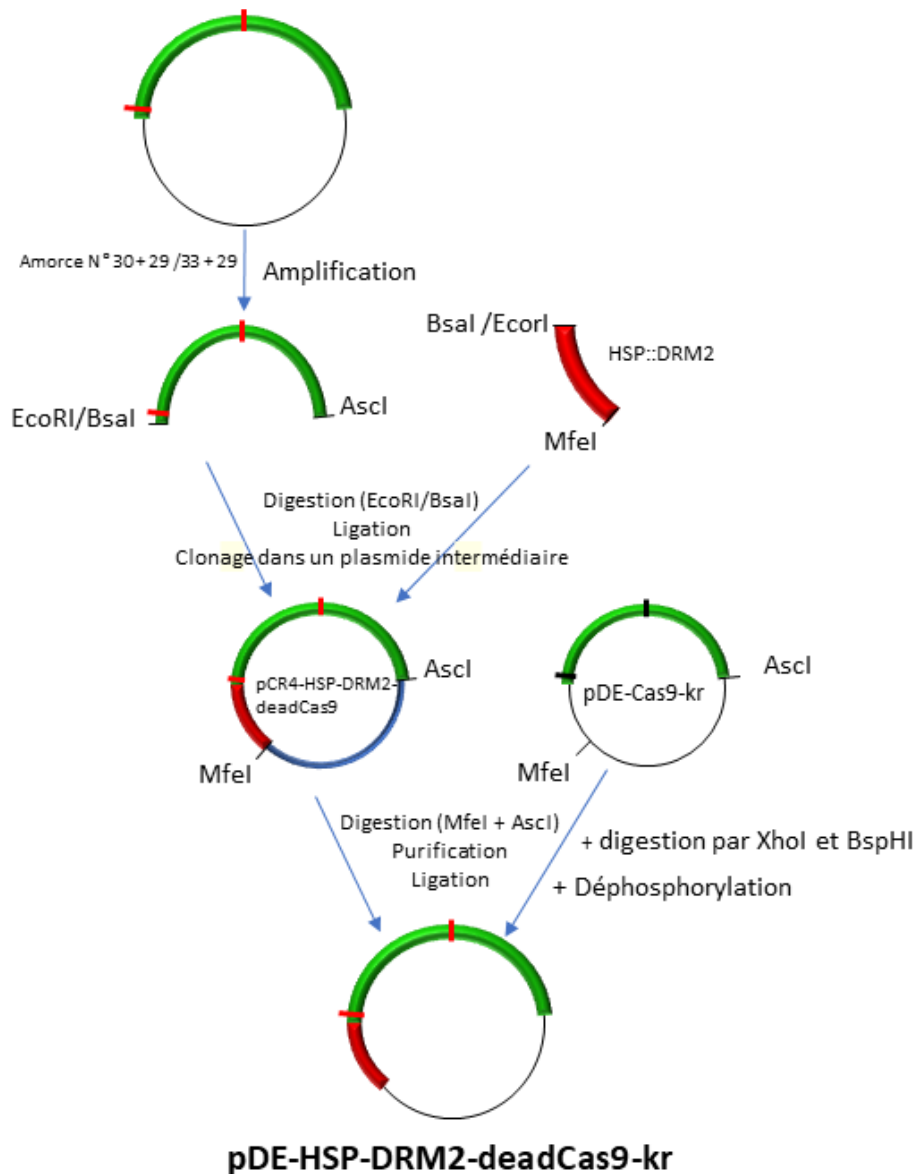


Figure 3.1-7: Schéma du protocole 5

d) Protocole d'insertion de la construction HSP::DRM2::deadCas9 synthétisée (Error! Reference source not found.)

Un plasmide de type pUC57 contenant la construction HSP::DRM2::deadCas9 a été commandé auprès de la société Protéogénix. La construction HSP::DRM2::deadCas9 est sortie du plasmide pUC- HSP-DRM2-deadCas9 par digestion avec les enzymes *EcoRI* et *AscI*. Le plasmide pDE-Cas9-kr est digéré par l'enzyme *AseI* et les enzymes *XhoI* et *BspHI* sont également utilisées pour digérer la Cas9 en 6 fragments, pour empêcher sa réinsertion dans le plasmide. Le fragment HSP::DRM2::deadCas9 et le plasmide pDE ouvert sont sélectionnés sur gel d'agarose et purifiés. Une digestion des extrémités cohésives via le kit « Quick Blunting™ Kit » (NEB) est réalisée pour obtenir des extrémités franches. Le plasmide pDE ouvert est ensuite déphosphorylé. Une ligation entre le fragment HSP::DRM2::deadCas9 et le plasmide pDE est ensuite réalisé à 10°C pendant 48h.

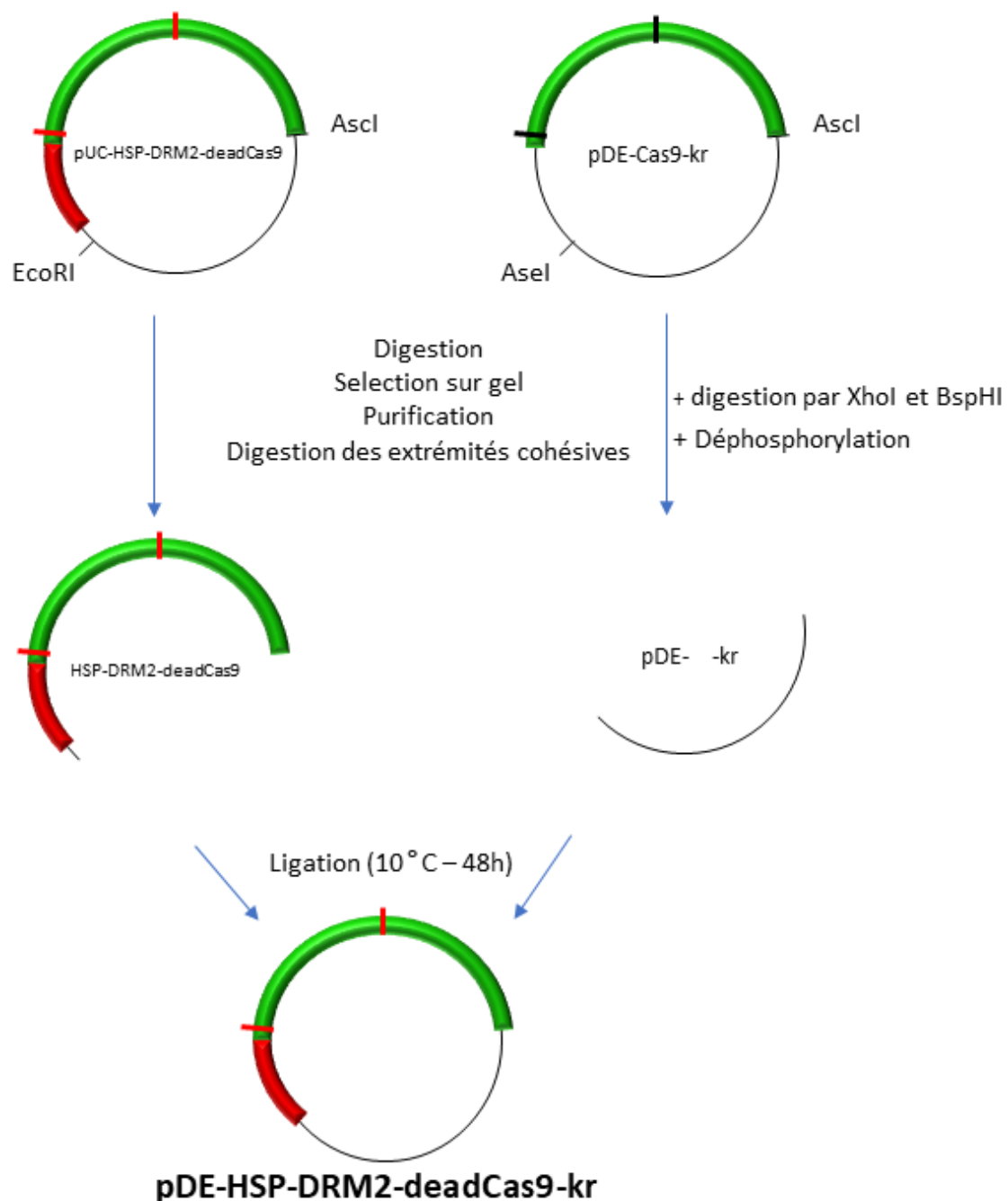


Figure 3.1-8 : Schéma du protocole 6

3.2. Résultats et discussion

3.2.1. Obtention des crRNAs ciblant le promoteur de *MdTFL1*

J'ai dans un premier temps observé la méthylation du gène *MdTFL1* sur les données de séquençage bisulfite obtenues précédemment (cf : chapitre I et Perrin et al 2020). On peut observer (**Error! Reference source not found.**A) que quel que soit l'état de maturité de la plante, la méthylation du promoteur semble équivalente. J'observe trois régions méthylées dans la région promotrice. La première région proche de la région 5'UTR du gène, porte principalement des marques de méthylation en contexte CG. Ensuite deux régions successives sont observables à plus de 500pb en amont du site d'initiation de la transcription. Ces régions sont riches en méthylation en contexte CHH. D'après les observations sur le niveau de transcription d'un gène en fonction de la distance

de la méthylation par rapport au site d'initiation de la transcription faites par Nicolas Daccord (Daccord, 2018), nous avons choisi de cibler une région de 500pb en amont du site d'initiation de la transcription.

J'ai donc réalisé une analyse des crRNAs possibles à l'aide du logiciel CRISPOR-Tefor précédemment présenté. Après sélection en fonction des différents critères de qualité (spécificité, présence de site de restriction, nombre de séquences d'appariement alternatives, etc...) il restait douze crRNAs possibles (**Error! Reference source not found.B**). J'ai ensuite choisi les crRNAs en les espaçant le plus possible les uns des autres pour éviter un éventuel problème d'encombrement stérique entre deux complexes CRISPR-deadCas9 qui se fixeraient simultanément. J'ai également exclu les crRNAs pouvant s'hybrider sur une région déjà riche en méthylation. J'ai ainsi sélectionné quatre crRNAs (en bleu clair sur la **Error! Reference source not found.B**).

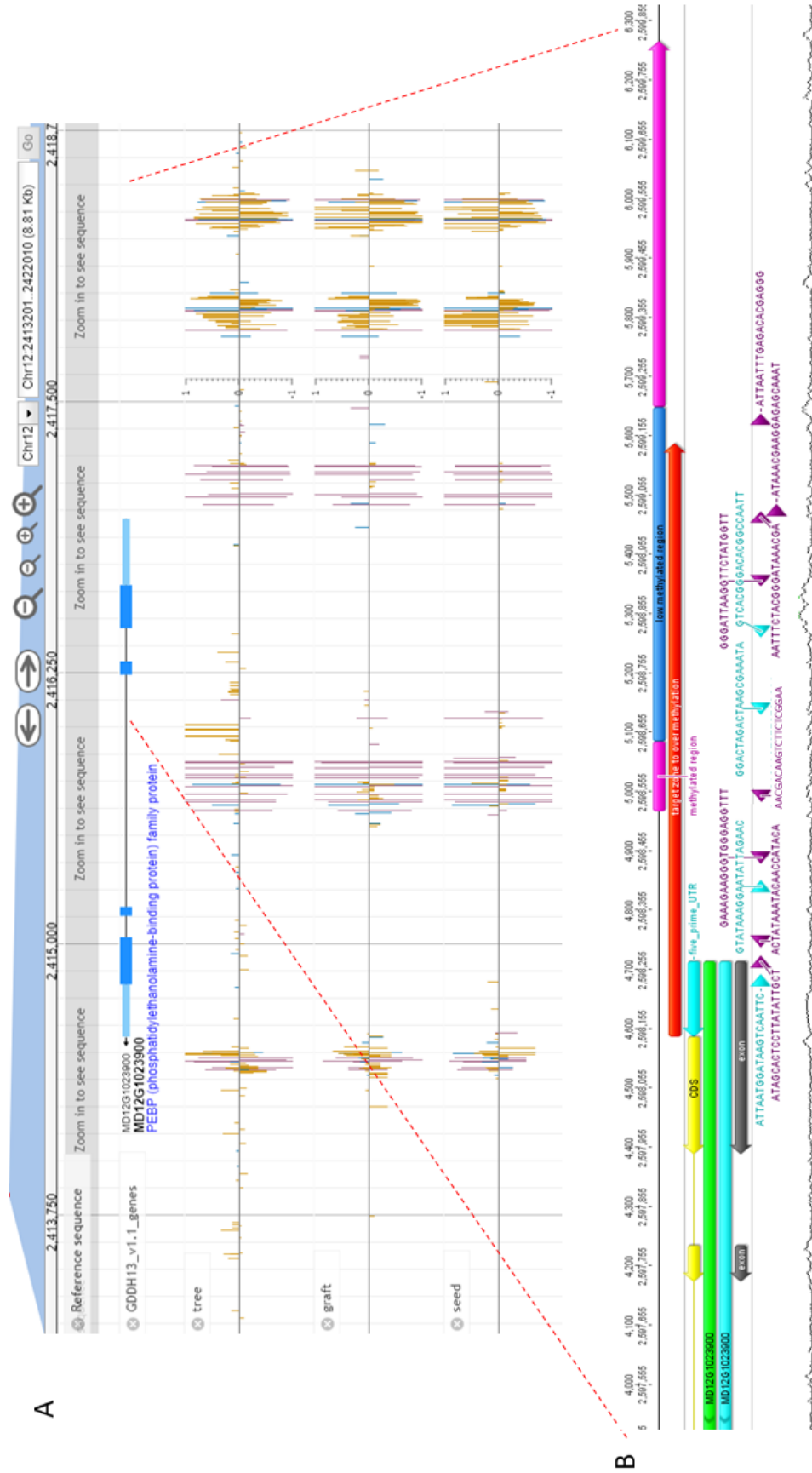


Figure 3.2-1 : Visualisation du gène *MdTFL1*. A. Capture réalisée sur le JBrowse local au locus du gène *MdTFL1*, avec les profils de méthylation de trois échantillons issus de la publication Perrin et al 2020. B. Positionnement des différents crRNA identifiés sur CRISPOR-Tefor dans la région promotrice du gène *MdTFL1*. Les triangles bleu clair représentent la position des crRNA sélectionnés, ceux en violet, d’autres crRNA possibles. Les rectangles roses indiquent les portions identifiées comme étant très méthylées dans la région promotrice du gène. Le rectangle bleu foncé indique une région identifiée comme moins méthylée.

3.2.2. Obtention d'une deadCas9 fonctionnelle

Pour réaliser les constructions précédentes (partie II-2) nous avons utilisé un plasmide provenant du laboratoire d'Olger Puchta (Fauser, Schiml, and Puchta (2014), Botanisches Institut, KIT, Allemagne). Nous avons donc choisi d'utiliser une Cas9 inactivée provenant également de leur laboratoire et obtenue à partir du plasmide pDE-Cas9 (Fauser *et al.*, 2014).

J'ai effectué des séquençages de PCRs (amorces N°34 + 35; 36 + 37; 38 + 39; 40 + 29, annexe 6) du plasmide pDe-deadCAS9-D10A-H840A pour vérifier la présence de la substitution d'acides aminés aux sites portant les activités nucléases. J'ai retrouvé les substitutions prévues, à savoir la substitution de l'acide aminé acide aspartique en alanine en position dix de la séquence protéique et la substitution d'une histidine en alanine en position huit-cents quarante de la séquence protéique. Mais j'ai constaté la présence de sept autres mutations qui n'étaient pas attendues dans la séquence nucléotidique, provoquant toutes une substitution d'acide aminé (Annexe 12: séquence envoyée par la laboratoire de Olger Puchta, non publié).

Nous avons donc décidé de repartir du plasmide pDe-Cas9-kr en notre possession, comportant une Cas9 active, et de réaliser nous-même les mutations permettant les modifications d'acides aminés en position 10 et 840 de la séquence protéique, pour obtenir le plasmide pDE-deadCas9-kr. Plusieurs stratégies s'offraient à nous : soit une invalidation de la Cas9 directement sur le plasmide, soit une invalidation de la Cas9 via des amplifications et clonages intermédiaires.

a) Premier protocole : mutation directement sur le plasmide pDeCas9-Kr (*Error! Reference source not found.*)

Ce premier protocole a le bénéfice d'être très simple à mettre en œuvre, car il suffit de créer autant de couples d'amorces que de mutations désirées dans le plasmide. Cependant je n'ai pas réussi à obtenir de résultat. L'amplification semble fonctionner, mais après la digestion par DpnI, puis la transformation, je n'ai observé aucun développement de colonie.

b) Second protocole : mutation par amplifications et clonages successifs à partir du plasmide pDeCas9-Kr (*Error! Reference source not found.*)

Le second protocole mis en place permet l'obtention d'une construction en cinq étapes. Ce protocole est plus long à mettre en place et à réaliser mais aurait dû permettre l'obtention de plasmides mutés plus facilement car nous réalisons des clonages successifs avec des vérifications à chaque étape. Cependant, ce protocole ne m'a pas permis d'obtenir de colonie positive pour l'insertion du plasmide pDedeadCas9-kr.

Initialement j'avais analysé les séquences à l'aide du logiciel Geneious v8.1.9 (<https://www.geneious.com>). Des analyses ultérieures de la séquence de la Cas9 avec un logiciel de prédiction des sites de restriction enzymatique différent (<http://www.restrictionmapper.org>) m'ont permis de trouver un site *BsaI* non initialement observé au milieu de la séquence de la Cas9. Ceci engendrait la création de deux fragments supplémentaires comportant des extrémités cohésives non complémentaires qui empêchaient donc la ligation des fragments entre eux et ainsi l'obtention d'un plasmide pDEdeadCas9-kr.

c) Troisième protocole : Utilisation d'une deadCas9 d'un plasmide existant (Error! Reference source not found.)

Lorsque j'ai fait part au laboratoire d'Olger Puchta de la présence de sites de mutations inattendus dans la séquence de la deadCas9 qu'ils m'avaient fait parvenir, ils ont retravaillé la séquence pour obtenir une deadCas9 présentant une séquence correcte, que j'ai obtenue. Cette deadCas9 était dans un plasmide différent de mon plasmide de destination. J'ai dans un premier temps séquencé cette deadCas9 avec les mêmes amorces (amorces N°34 + 35; 36 + 37; 38 + 39; 40 + 29, annexe 6). Je n'ai pas retrouvé de mutations à part celles souhaitées, la séquence correspondait bien à la séquence attendue de la deadCas9.

Par la suite, via les sites *AscI* présents aux extrémités de la séquence, j'ai pu remplacer la Cas9 active du plasmide pDE-Cas9-kr par cette deadCas9. J'ai ensuite vérifié l'intégration de la deadCas9 par PCR et séquençage.

J'ai donc pu obtenir le plasmide pDE-deadCas9-kr contenant le gène de sélection à la kanamycine ainsi que les sites de restriction utiles pour la suite.

3.2.3. Echec de l'obtention de la construction HSP::DRM2 ::deadCas9 dans le plasmide final

Lors de la préparation des clonages *in silico* nous avons prévu d'insérer le fragment HSP::DRM2 directement dans le plasmide pDE-deadCas9-kr précédemment obtenu. Pour cela nous devions utiliser les enzymes *BsaI* ou *EcoRI* et *MfeI* pour ouvrir le plasmide et réaliser l'insertion dans le plasmide sans étapes intermédiaires. Cependant nous avons initialement mal choisis nos enzymes pour réaliser cette étape : *BsaI*, *EcoRI* et *MfeI* coupent toutes le plasmide à d'autres sites que ceux souhaités (annexe 13 et **Error! Reference source not found.D**). Ces erreurs nous ont obligé à réamplifier la deadCas9 à partir du plasmide pDE-deadCas9-Kr pour tenter d'obtenir la construction HSP::DRM2::deadCas9 complète tout d'abord, puis son insertion dans le pDE-Cas9-Kr en remplacement de la Cas9 active.

a) Choix de la méthyltransférase DRM2

Pour réaliser cette partie de la construction nous nous sommes inspirés de ce qui avait été réalisé sur cellules de mammifère. Liu et al. (2016) et Vojta et al. (2016) ont tous deux utilisé l'activité de la méthyltransférase humaine DNMT3a pour éditer l'épigénome. Chez les plantes l'orthologue de *Dnmt3a* est le gène *Drm2*, il permet une méthylation *de novo* (cf. Introduction partie 4.3.3 a)) et peut cibler les cytosines dans les trois contextes (CG, CHG, CHH) via la RdDM (cf. Introduction partie 4.3.3 b), Law and Jacobsen (2010)).

Les séquences choisies ont été intégrées *in silico* dans des constructions nommées HSP::DRM2S, T ou KO, que nous avons fait synthétiser par l'entreprise Protéogénix.

b) Obtention de la construction HSP::DRM2 ::deadCas9 par Golden Gate (Error! Reference source not found.)

Ce protocole devait permettre d'obtenir de façon simple et rapide la construction HSP::DRM2 ::deadCas9 en utilisant la technique du Golden Gate. Mais il ne pouvait en fait pas fonctionner car, comme vu aussi en 3.2.2 b) de cette partie, le site *BsaI* non initialement remarqué au milieu de la Cas9 (comme de la deadCas9) empêchait que la réaction se fasse correctement.

c) Obtention de la construction HSP::DRM2 ::deadCas9 par clonage conventionnel (*Error! Reference source not found.*)

Nous avons également la possibilité de réaliser un clonage conventionnel en utilisant l'enzyme *EcoRI*. Cependant je n'ai pas réussi à obtenir de colonies présentant l'insertion de la construction HSP::DRM2::deadCas9 dans le plasmide final via ce protocole.

d) Protocole d'obtention à l'aide d'un plasmide intermédiaire (*Error! Reference source not found.*)

J'ai également essayé de passer par un plasmide spécialement conçu pour les gros fragments (zero blunt TOPO®, thermofisher), pour réaliser un clonage intermédiaire de la ligation HSP::DRM2::deadCas9, avant insertion dans le plasmide final. Comme vu précédemment, le site *BsaI* non initialement identifié au milieu de la Cas9/deadCas9 invalidait ce protocole quand associé au Golden Gate. Associé au clonage conventionnel via digestion/ligation aux sites *EcoRI*, ce protocole ne m'a pas non plus permis d'obtenir de bactéries positives pour le plasmide pCR4-HSP-DRM2-deadCas9.

J'ai finalement refait une annotation des sites de restriction présents sur les séquences, avec un autre logiciel que celui initialement utilisé. J'ai découvert la présence d'un site de restriction *MfeI* non initialement observé dans les séquences HSP::DRM2 (T, S ou KO) synthétisées. Cette digestion supplémentaire diminuait les chances d'obtenir une ligation complète car elle augmentait le nombre de fragments à religuer ensuite. Elle pouvait aussi aboutir à des ligations non souhaitées avec l'insertion du fragment tronqué.

e) Protocole d'insertion de la construction HSP::DRM2::deadCas9 synthétisée (*Error! Reference source not found.*)

Etant donné les problèmes rencontrés précédemment, nous avons finalement commandé la synthèse de la construction HSP::DRM2::deadCas9. Avec l'aide de Jaiana Malabarba (post-doctorante à l'IRHS, ResPom), nous avons mis en place un protocole différent utilisant d'une part les enzymes *AscI* et *EcoRI* pour récupérer le fragment HSP::DRM2::deadCas9 dans le plasmide commandé (pUC57HSP ::DMR2 ::deadCas9) et d'autre part l'enzyme *AscI* (deux sites de part et d'autre de la Cas9) pour ouvrir et éliminer la Cas9 active du plasmide pDE-Cas9-kr. Nous avons obtenu des colonies positives d'après les résultats PCR. Mais après multiplication, les résultats PCR et de séquençage étaient systématiquement négatifs.

Ce problème semble assez récurrent pour les clonages considérés comme « difficiles », sans qu'une explication soit pour l'instant avancée (Forum Research Gate, 2013).

Lors de nos essais de clonage, nous avons également effectué des PCRs sur le fragment HSP ::DMR2 ::deadCas9 que nous avons fait synthétiser. Nous avons utilisé différents couples d'amorces, mais le couple permettant d'amplifier la ligation entre DRM2 et deadCas9 (amorces n° 44 et 45, annexe 6) au niveau des GS linker ne fonctionnait pas bien. Les fragments obtenus avaient des tailles aberrantes et leur séquençage ne donnait pas la séquence attendue. Il semble que cette séquence répétée de GS linker pose donc également un problème dans la structure.

3.2.4. Proposition d'un protocole amélioré pour l'obtention de la construction HSP::DRM2::deadCas9 dans le plasmide final

a) Synthèse complète

Afin d'obtenir un plasmide contenant la construction HSP::DRM2::deadCas9, une solution serait de faire synthétiser complètement le plasmide. Nous pourrions dans ce cas choisir un autre « squelette » de plasmide, plus petit que celui du pDE, tout en conservant la cassette de sélection par la kanamycine (importante pour la sélection *in vitro* chez le pommier) ainsi que les sites AttR nécessaires pour réaliser une réaction de Gateway® afin d'intégrer les gRNAs.

Un plasmide plus petit faciliterait la synthèse et diminuerait les coûts. En effet, le plasmide actuel nécessiterait une synthèse de plus de 16kb pour le plus petit (construction avec DRM2S et DRM2KO) et de plus de 18kb pour le plus grand (construction avec DRM2T). J'ai contacté la société Proteogenix qui m'a répondu ne pas être en mesure de réaliser une telle synthèse.

b) Protocole à mettre en place pour obtenir le plasmide pDE-HSP-DRM2-deadCas9-kr

Afin d'obtenir le plasmide pDE-HSP-DRM2-deadCas9-kr, j'ai réétudié la séquence du plasmide pDE-deadCas9-Kr, notamment pour identifier les enzymes de restrictions utilisables dans ce plasmide. Je propose donc le protocole suivant (**Error! Reference source not found.**).

- (1) La deadCas9 est amplifiée avec les amorces N°41 et 29 (annexe 6). L'amorce N°41 permet d'ajouter deux sites de restriction : *ApaI* et *SwaI*. L'amplification est réalisée avec la polymérase haute-fidélité « Q5 » (NEB).
- (2) Le plasmide pDE-deadCas9-kr est digéré par *AscI* pour retirer la deadCas9. Une étape de digestion des extrémités cohésives est réalisée.
- (3) Une ligation est réalisée entre le fragment de la deadCas9 amplifié et le plasmide ouvert. Après transformation, une sélection par PCR et séquençage est réalisée pour ne conserver que les plasmides ayant intégrés la deadCas9 amplifiée avec les sites *ApaI* et *SwaI*. Le plasmide est ensuite digéré avec *ApaI* et *SwaI* et déphosphorylé pour l'étape (5).
- (4) Les constructions HSP-DRM2 (S, T et KO) sont amplifiées avec les amorces N°42 et 43 par la polymérase haute-fidélité « Q5 » (NEB). L'amorce N°42 permet d'ajouter un site *SwaI* en 5' des constructions et l'amorce N°43 permet d'ajouter un site *ApaI* en 3' des constructions (annexe 6) et de supprimer les séquences « GS ». Chaque amplicon est ensuite digéré avec les enzymes *ApaI* et *SwaI*.
- (5) La ligation du plasmide pDE-dCas9-kr ouvert et des différentes constructions issues de (4) est réalisée à température ambiante (20°C) pendant une nuit, en utilisant différentes concentrations d'inserts et de plasmide.
- (6) La ligation est ensuite utilisée pour transformer des bactéries résistantes au « ccdB » et électro-compétente afin d'optimiser la transformation. Puis des vérifications par séquençage sont réalisées.

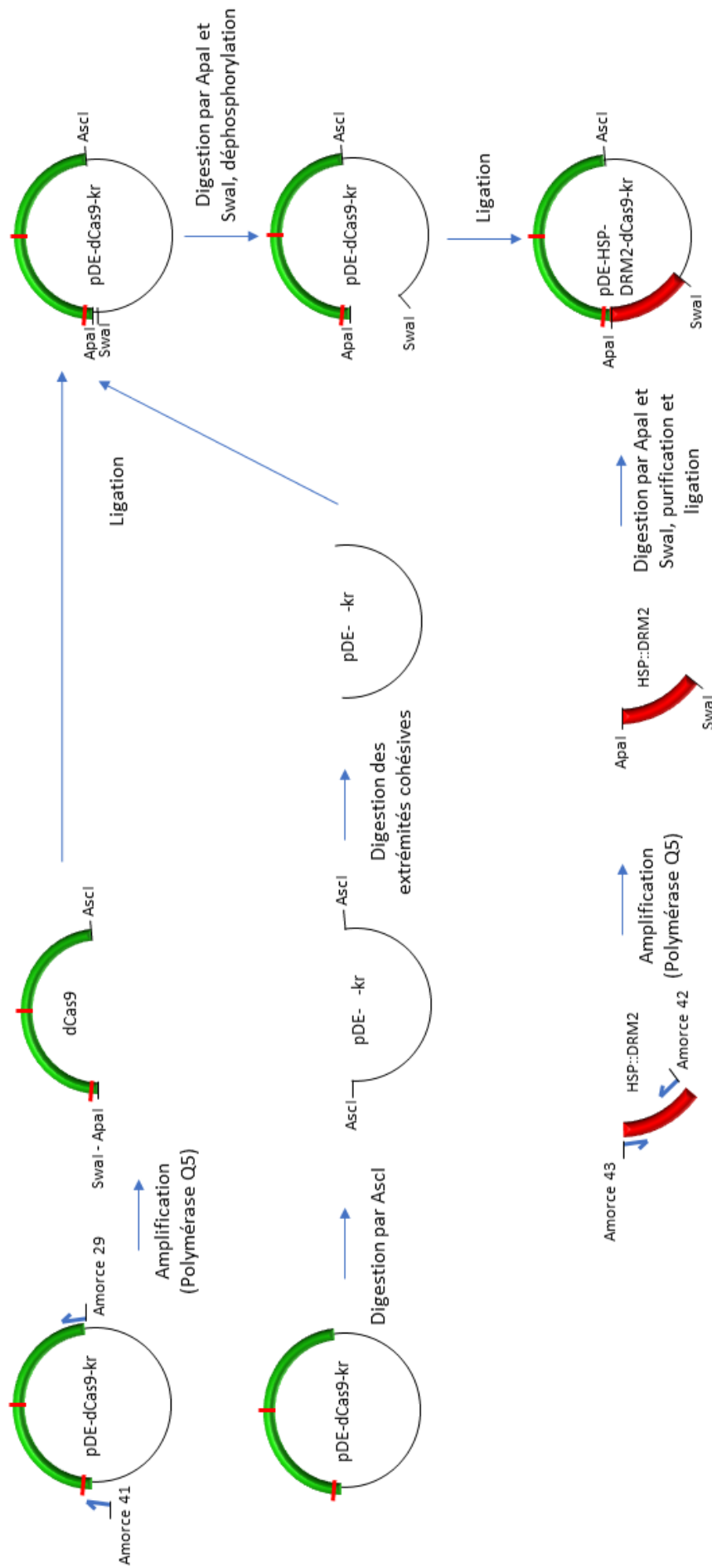


Figure 3.2-2: Proposition de protocole

4. Conclusion – Chapitre II

Tout d'abord concernant la partie 3, l'hypothèse de travail était qu'en hyper-méthylant le promoteur du gène codant le répresseur floral *MdTFL1* (*MD12G1023900*), nous allions réduire sa transcription, et ainsi obtenir des plantes fleurissant précocement, après application d'un stress thermique.

Nous n'avons pas réussi à obtenir la construction nécessaire à ce travail dans le temps imparti de ma thèse. Il serait cependant intéressant de poursuivre les efforts engagés. En effet, obtenir des pommiers fleurissant à la demande, donc plus précocement que de manière naturelle, permettrait de réaliser des croisements plus rapidement et donc de réduire le temps de sélection des nouvelles variétés. Ce type de lignée à floraison précoce ou à floraison inductible a été obtenue par d'autre méthode que la technique CRISPR et est déjà utilisée dans des programmes de sélection rapide (« Fast-breeding », Schlathölter et al. (2018); Wenzel, Flachowsky, and Hanke (2013)).

Concernant la partie 2, j'ai obtenu des plantes éditées pour l'un des orthologues de *AtMET1* chez le pommier : *Md15MET1*. Le travail que nous avons réalisé jusqu'à présent ne permet pas encore d'envisager une publication mais il est poursuivi. Les paragraphes suivants détaillent ce qu'il nous paraît encore important de réaliser en vue d'une publication.

Dans un premier temps nous devons réaliser des mutants du second gène, *Md02MET1*. En effet, nous ne savons pas si le gène *Md15MET1* est le seul impliqué dans le phénotype de maintenance de la méthylation en contexte symétrique CG chez le pommier et les essais réalisés en Q-PCR pour observer le niveau de transcription de ces gènes ne sont pas concluants pour l'instant.

Il serait également intéressant de tenter d'obtenir des doubles mutants pour les gènes *Md15MET1* et *Md02MET1*. Dans le cas où les transformations correspondantes seraient des échecs, mais que les témoins de transformations associés aboutiraient eux à l'obtention de bourgeons transformés, une hypothèse pourrait être que la double mutation est létale.

Il faudrait également réaliser une étude comparative de la transcription des gènes *Md15MET1* et *Md02MET1*, notamment dans des organes à forte multiplication cellulaire (par exemple sur cal régénérant et sur bourgeons) étant donné la fonction du gène. Cette étude permettrait en particulier de définir dans quelles parties de la plante ces gènes s'expriment, et dans quelle mesure. On notera que He et al. (2017) ont mesuré l'expression du gène *MET1* dans différents organes de la variété 'Hanfu' (*Malus domestica*). Ils ne semblent cependant pas avoir dissocié les deux orthologues de *MET1*.

Par ailleurs, l'étude de l'expression du gène *Md02MET1* chez les plants édités pour *Md15MET1* permettrait d'évaluer dans quelle mesure l'activité de ces deux gènes est redondante. Une étude similaire a déjà été réalisée chez *M.domestica*, pour les gènes *MdbHLH001* and *MdbHLH002* codant pour une protéine de type « Basic helix-loop-helix » (Yang et al., 2017), montrant que lorsque l'un des deux gènes est muté, le second assure une activité suffisante pour accomplir la fonction. Une telle étude nécessite cependant d'obtenir un mutant du second homologue (*Md02MET1*) et un double mutant.

Afin de confirmer l'absence globale (excepté pour une plante dans une lignée) de modifications morphologiques des lignées, observée à l'œil nu, nous devons également obtenir des données de

phénotypage au niveau macroscopique et microscopique. Il serait ainsi opportun de (i) mesurer la vitesse de croissance et la forme des feuilles et leur densité de trichomes, (ii) doser leur teneur en matière sèche, (iii) mesurer la taille des entre nœuds, et (iv) évaluer le retard de floraison. Il a été montré que la méthylation du gène *AtFWA* est contrôlé par *MET1* et qu'une perte de la méthylation sur ce gène induit une floraison tardive chez *A.thaliana* (Kinoshita, 2004). Ces caractères ont été rapportés comme modifiés chez *A.thaliana* par l'inactivation de *MET1*, par l'insertion d'une construction en antisens (Ronemus *et al.*, 1996a; Kim *et al.*, 2008), ou par traitement EMS (Kankel *et al.*, 2002). Pour les caractères microscopiques il serait intéressant d'observer l'organisation et l'épaisseur du limbe sur coupes transversales, ainsi que de la nervure centrale, afin de détecter d'éventuels effets sur les tissus vasculaires.

Enfin l'épigénotypage d'une partie des lignées reste à effectuer. Dans un premier temps il faudrait observer s'il y a une diminution de la méthylation en contexte CG, globale ou localisée. En cas de diminution localisée, il faudrait alors, par le calcul de DMRs, analyser si des voies particulières sont plus ciblées que d'autres. Une analyse du mobilome de ces mutants seraient également à envisager, afin d'évaluer la mobilité des transposons induite par une diminution de la méthylation.

Une étude sera également menée par l'équipe VALEMA de l'IRHS (Jean-Marc Celton), sur les lignées 2C3 et 2H1, éditées pour le gène *Md15MET1*. L'objectif sera de tester la tolérance à un stress hydrique dans différentes conditions (*in vitro*, ou en serre). Les mutants pour la fonction *MET1* seront testés *in vitro* et en serre.

Enfin les mutants *Mdmet1* obtenus pourraient permettre d'approfondir la compréhension de la transmission intergénérationnelle des marques épigénétiques associées à la voie de la maintenance de la méthylation. En particulier, l'étude de la transmission via la multiplication asexuée (greffage) pourrait être réalisée rapidement. Cependant l'étude de la transmission via la multiplication sexuée ne pourra pas être réalisée avec ces mutants car ils ont pour fond génétique le génotype 'Gala', hautement hétérozygote et auto-incompatible. Une étape de reproduction sexuée nécessiterait donc que le second parent soit un autre génotype et les descendants seraient tous différents (génétiquement et donc phénotypiquement) des parents, du fait du brassage génétique. Or pour réaliser une telle étude, il est nécessaire que le génotype reste identique, afin de pouvoir éliminer la variation génétique, plutôt que les modifications épigénétiques, comme source de différence. Une étude similaire à celle réalisée dans Perrin et al 2020 ne sera donc pas possible tant que des mutants pour la fonction *MET1* n'auront pas été obtenus chez le génotype homozygote et auto-fertile GDDH13.

Conclusion générale

Dans le cadre de ce travail de thèse j'ai dans un premier temps cherché à comprendre quels sont les conséquences épigénétiques de la multiplication d'une plante pérenne telle que le pommier par propagation asexuée par rapport à une multiplication sexuée.

Du point de vue épigénétique, il est important de bien séparer la reproduction sexuée, qui permet au genre *Malus* de se reproduire à l'état sauvage et qui permet également aux sélectionneurs de réaliser des croisements pour obtenir de nouvelles variétés, de la propagation asexuée réalisée par greffage qui est utilisée pour la production de fruits.

En ce qui concerne la propagation asexuée, la production de pommiers pour les vergers repose sur la propagation par greffage. Cette propagation permet en premier lieu une conservation de l'homogénéité des variétés inscrites au « catalogue officiel des espèces et variétés de plantes cultivées en France », puisque le pommier présente une auto-incompatibilité pollinique (Zhang *et al.*, 1988). En second lieu, le greffage permet de réduire la période juvénile du pommier, de 8 ans minimum chez un pommier issu de semis (Visser, 1964) à deux ou trois ans entre l'obtention du plant greffé et son entrée en fructification (Simons, 1986).

La propagation clonale a été largement utilisée pour la domestication des plantes cultivées car elle permet notamment de sélectionner et de conserver des caractères agronomiques évalués comme utiles (discuté dans McKey *et al.*, (2010)). L'utilisation du greffage par les multiplicateurs permet la propagation rapide d'un génotype favorable (McKey *et al.*, (2010)). Le greffage chez une espèce hautement hétérozygote telle que le pommier, permet d'éviter la dépression de sélection existant lors d'une fécondation croisée (Noël *et al.*, 2017). Les multiplicateurs ont également privilégié cette technique pour conserver les propriétés organoleptiques des fruits au sein d'une même variété (McKey *et al.*, (2010)).

Pendant ce travail de thèse j'ai étudié l'impact de la multiplication d'une variété de pomme dérivée de 'Golden Delicious', par propagation asexuée par rapport à une multiplication sexuée, sur son épigénome et son phénotype.

En effet, j'ai pu mettre en évidence des caractères phénotypiques reflétant la différence entre plantes juvéniles et plantes adultes matures tel que des feuilles présentant une densité de trichomes plus élevée chez des plantes âgées. Cette densité de trichomes pourrait être associée à la défense contre les herbivores. Une plante juvénile développant des trichomes progressivement suite à des attaques d'herbivores (Dalin *et al.*, 2008). Roda *et al.*, (2003) ont émis l'hypothèse que cette plus grande densité de trichomes sur les feuilles, permettrait d'augmenter la capture de pollen et de spores de champignon. Ceux-ci serviraient de source de nourriture alternative aux herbivores situés sur les feuilles, limitant ainsi les blessures faites aux feuilles. J'ai également observé que les plants issus de greffage présentaient un état intermédiaire que l'on peut décrire comme adulte immature. En effet, ces plants présentaient une densité de trichomes similaire aux plantes adultes matures, mais n'étaient pas capable de fleurir. Nous avons conclu que cet état d'adulte immature était lié à une différence de gradient auxinique défini par la distance entre le porte greffé et l'apex (Zimmerman, 1973; Hanke *et al.*, 2007; Zhang *et al.*, 2007).

Dans un second temps j'ai analysé le niveau de transcription des gènes au sein de ces plants de différents âges et issus de différents modes de propagation. J'ai constaté une différence de transcription graduelle : les semis présentant un taux de transcription globalement plus élevé que les plants greffés. Et les arbres adultes matures présentant un taux de transcription globalement plus faible que les semis et les greffons. Ceci est notamment associé à une modification de l'activité photosynthétique (Hutchison *et al.*, 1990; Greenwood, 1995; Bond, 2000).

Enfin j'ai réalisé une analyse des méthylomes de ces différentes plantes que j'ai ensuite comparés. J'ai ainsi pu mettre en évidence que la variété 'GDDH13' possède une empreinte épigénétique transmise de façon relativement stable que ce soit à travers la multiplication sexuée ou la propagation asexuée (Perrin *et al.*, 2020). J'ai cependant identifié des régions différentiellement méthylées correspondant à environ 3% du génome pour la multiplication sexuée et 1,35% du génome pour la propagation asexuée de notre modèle d'étude. En 2002, Fraga *et al.*, ont montré chez *Pinus radiata* que des étapes de greffage successives tendent à diminuer progressivement le taux de méthylation. Ainsi après quatre cycles de greffage, le taux de méthylation d'un greffon issu d'un arbre mature se rapprochait de celui d'un arbre juvénile. Nous avons noté une différence de méthylation plus faible entre les greffons issus d'un cycle de greffe unique et l'arbre adulte qu'entre ce dernier et les plants issus de semis. Ce qui semble cohérent avec les observations de Fraga *et al.*, (2002). Il semble donc que la multiplication qu'elle soit sexuée ou asexuée implique des modifications locales de l'empreinte épigénétique (les DMRs), mais ne modifie pas l'empreinte épigénétique « basale » de l'espèce.

En outre l'observation de régions différentiellement méthylées entre les arbres adultes matures et les greffons pourrait être dû à la combinaison de deux stress. Un stress dû à la blessure causée par le prélèvement des scions sur l'arbre parent et un stress causé par le greffage en lui-même (Cookson *et al.*, 2013; Cookson & Ollat, 2013). Puisque que le porte greffe et le scion n'ont pas le même génome il doit y avoir une forme d'acclimatation de l'un et de l'autre, afin d'éviter ce que l'on pourrait appeler un rejet de la greffe. Cette acclimatation pouvant être lié notamment à la modification de l'épigénome des deux partis via l'échange de petits ARNs (Flachowsky *et al.*, 2012). Les arbres adultes matures sont également greffés, cependant le stress lié au greffage à eu lieu il y plusieurs dizaines d'années, leur épigénome a donc probablement eu le temps de se stabiliser et « d'oublier » ce stress.

Les modifications épigénétiques observées entre les greffons et les arbres adultes matures pourraient également être le reflet de modifications physiologiques tels qu'une modification du gradient auxinique (Zhao, 2010). Entre un jeune greffon et un arbre adulte, le gradient auxinique n'est pas le même puisque la distance entre le porte greffe et l'apex est différente. Or nous avons prélevé des feuilles tout juste formées au niveau de l'apex, donc à une distance différente du porte greffe entre les greffons et l'arbre adulte. De plus il a été montré que les voies de biosynthèse et de transport de l'auxine sont en partie régulées par des mécanismes épigénétiques (Li *et al.*, 2011; Mateo-Bonmatí *et al.*, 2019). En outre, Corredoira *et al.*, (2017) laisse supposer un lien entre auxine et méthylation de l'ADN durant l'embryogénèse somatique. Il est donc probable qu'une partie des DMRs observées entre les greffons et l'arbre adulte soit associée à la modification du gradient auxinique au cours du processus de greffage. Ces DMRs ne seraient donc qu'indirectement liées à la propagation asexuée. Afin d'identifier ce qui est lié à la propagation asexuée et ce qui est lié à la

blessure causée par le greffage, il sera nécessaire de réaliser des autogreffes sur des plants issus de semis afin de comparer l'évolution de leur méthylome par rapport à celui des semis.

Lors de la multiplication sexuée il doit exister un mécanisme d'adaptation rapide permettant à la plante de supprimer ou de restaurer les marques de méthylation ayant été modifiées dans les plantes parents. Nous supposons donc que la conservation de ces modifications aurait un coût énergétique ou développemental trop important pour la plante. C'est notamment ce que définissent Hollister & Gaut, (2009) dans leur analyse de la répression épigénétique des TEs chez *A. thaliana*. Au sein des cellules, il se met en place un compromis entre répression et non répression des TEs, affectant l'expression des gènes dans leur voisinage. Ainsi pour une jeune plante issue de multiplication sexuée la priorité n'est pas la défense contre les pathogènes ou les herbivores. Cette défense pourrait être favorisée par la mémoire d'un stress biotique vécu par les arbres parents, mais aurait un coût énergétique trop important (Dalin *et al.*, 2008). Il est plus important pour cette plantule de concentrer ses ressources énergétiques sur l'accès à la lumière et de mettre en place les processus de photosynthèse et d'établissement de son système racinaire. Une fois établie la plante pourra ensuite réorienter l'expression de ses gènes pour s'adapter aux contraintes biotiques et abiotiques existantes dans son environnement. La suppression des modifications épigénétiques acquises par les arbres parents permettrait donc à la plantule issue de multiplication sexuée de restaurer une « valeur sélective » (ou « fitness » en anglais) permettant une meilleure survie lors des étapes de développement précoce. Il est également envisageable que pour la jeune plante, il soit plus facile de se créer sa propre mémoire épigénétique afin de s'adapter à ses contraintes environnementales, plutôt que d'utiliser une mémoire héritée de ses parents qui ne serait pas forcément optimale.

Dans le cas de la propagation asexuée par greffage, nous avons également observé une suppression des modifications épigénétiques, mais moindre que pour la multiplication sexuée. Il est ici important de noter que les pommiers greffés ne sont pas soumis aux mêmes contraintes que les pommiers issus de semis. Le greffage a un intérêt économique pour l'Homme. Durant ses premières étapes de développement le greffon n'est pas soumis à une compétition pour l'accès à la lumière. On peut également supposer que les premières étapes de développement des greffons seront plutôt orientées vers une réorganisation des tissus vasculaires, afin d'établir une communication stable et durable entre le porte-greffe et le scion (Kumari *et al.*, 2015) comme montré chez la vigne par Cookson *et al.*, (2013) et Moreno *et al.*, (2014). Les greffons étant plutôt cultivés en monoculture, on peut également supposer qu'ils font face à des contraintes biotiques plus importantes du fait de la monoculture (Newton *et al.*, 2010; Felton *et al.*, 2016; Wu *et al.*, 2020). Une hypothèse serait que la transmission des marques épigénétiques par greffage est plus importante pour les marques associées à l'adaptation aux contraintes biotiques et abiotiques dans les vergers. Supprimer la mémoire d'un stress biotique ou abiotique serait donc plus dommageable pour les greffons.

Par le biais des activités humaines de greffage nous avons sélectionné de nombreux caractères agronomiques intéressants pour l'Homme (Robinson, 2008; Melnyk & Meyerowitz, 2015). Nous pouvons également émettre l'hypothèse que via nos activités de greffage, nous avons artificiellement sélectionné une voie de transmission des marques épigénétiques chez les plantes cultivées issues de greffage. Cette voie de transmission diffère de la façon dont sont transmises les marques épigénétiques lors de la multiplication sexuée.

L'épigénétique est également un facteur d'adaptation des plantes aux contraintes biotique et abiotique (Richards *et al.*, 2010; Mirouze & Paszkowski, 2011; Casadesús & Low, 2013). Nous

pouvons supposer que la transmission des marques épigénétiques joue un rôle important dans l'adaptation des plantes à propagation asexuée, face aux contraintes biotiques et abiotiques, car elle offre une plasticité transcriptionnelle et phénotypique importante. Ainsi, depuis sa description il y a plus de cent ans (Company Stark Brothers, 1913), la variété 'Golden Delicious' a subi de très nombreux cycles de propagation et a donc potentiellement accumulé des mutations génétiques. On peut également supposer que son épigénome a évolué, d'autant plus que le taux de mutation épigénétique est supérieur ($\sim 10^{-4}$, Van der Graaf *et al.*, (2015)) au taux de mutation génétique ($\sim 10^{-8}$, Ossowski *et al.*, (2010) chez *A. thaliana*). Nous pouvons donc nous demander si cette variété est toujours la même aujourd'hui. Afin d'obtenir une partie de la réponse, il pourrait être envisagé d'étudier de vieux pommiers, conservés dans les arboretums, dans les collections de l'INRAe ou dans des collections privées. Sur ces vieux arbres des carottages pourront être réalisés dans le tronc afin d'effectuer un séquençage bisulfite des anneaux de croissance. Le méthylome des anneaux les plus anciens pourra ainsi être comparé aux plus récents pour observer l'évolution du méthylome dans le temps. Si ces arbres portent encore des fruits, il serait également intéressant de comparer leur composition chimique et organoleptique par rapport à des fruits issus d'un arbre jeune de la même variété.

Pendant ma thèse j'ai également expérimenté des techniques permettant l'obtention de mutants ou d'épi-mutants dans le cadre du projet Genius. Les premiers résultats obtenus pour cette partie pourront être valorisés par la suite afin de tester les différentes hypothèses discutées ci-dessus. En effet nous avons défini que les semis de pommiers sont dans un état naïf de méthylation par rapport à un arbre adulte mature et nous avons relié cet état à une augmentation de la transcription des gènes, notamment des gènes impliqués dans la photosynthèse (Perrin *et al.*, 2020). Il serait donc intéressant de réaliser des mutants ayant intégrés une construction inductible permettant de déméthylater ou de méthylater une cible spécifique dans le génome. Ces plants devront ensuite être greffés avant de réaliser une induction de la construction afin de déméthylater ou de méthylater spécifiquement les gènes impliqués dans la photosynthèse. Il sera ainsi intéressant d'analyser si la diminution de méthylation sur ces gènes permet d'augmenter la valeur sélective des greffons et d'induire un état naïf de méthylation similaire à celui des semis.

A l'avenir, il est envisageable de penser que l'amélioration variétale ne se fera plus uniquement sur des aspects génétiques mais qu'elle prendra en compte l'épigénétique comme un outil complémentaire. On peut notamment envisager de préparer les plantes en pépinières, via l'épigénétique, aux conditions pédoclimatiques auxquelles elles feront face sur leur lieu de culture. Cependant, comme nous l'avons noté dans les différentes comparaisons réalisées entre 2,64 et 6,55% des gènes annotés dans le génome sont différenciellement transcrit et présentent une modification locale du niveau de méthylation. Il semble donc que dans le génome du pommier peut de gènes sont régulés par la méthylation. (Daccord *et al.*, 2017) ont d'ailleurs montré qu'un haut niveau de méthylation, relevé sur environ 10% des gènes de GDDH13, semble corrélé au niveau de transcription d'un gène. L'utilisation de l'épigénétiques dans les programmes d'amélioration du pommier plus précisément sera donc restreinte aux gènes méthylés du pommier pour lesquelles la méthylation a un effet sur le niveau d'expression. Actuellement l'amélioration des plantes via l'épigénétiques se base sur la découverte d'épi-allèle naturelle ou induit via sélection (House & Lukens, 2019). Les marques épigénétiques sont très labiles au cours de la vie des plantes. (House & Lukens, 2019) ont d'ailleurs défini trois types d'épi-allèle dont deux semblent plus stables : les épi-allèles « faciliated » et obligatoire qui sont dépendants de la séquence ADN sous-jacente.

L'utilisation de l'épigénétique pour la sélection variétale via ce type d'épipallèle sur base sur l'induction de la RdDM via l'insertion d'une séquence homologue de la séquence cible afin d'induire une hyperméthylation de la séquence cible pour réduire la transcription des séquences associées.

Bibliographie

- Adams RL. 1990.** DNA methylation. The effect of minor bases on DNA-protein interactions. *The Biochemical journal* **265**: 309–20.
- Ahmed MJ, Murtaza Gillani SG, Ayaz Kiani F. 2013.** Response of apple (*Malus domestica* Borkh.) cultivars grafted on two rootstocks under sub-humid temperate climate of Azad Jammu and Kashmir. *Pakistan Journal of Agricultural Sciences* **50**: 379–386.
- Albaladejo RG, Parejo-Farnés C, Rubio-Pérez E, Nora S, Aparicio A. 2019.** Linking DNA methylation with performance in a woody plant species. *Tree Genetics and Genomes* **15**.
- Alberts B. 2003.** DNA replication and recombination. *Nature* **421**: 431–435.
- Almeida R, Allshire RC. 2005.** RNA silencing and genome regulation. *Trends in Cell Biology* **15**: 251–258.
- Anzalone A V., Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, Chen PJ, Wilson C, Newby GA, Raguram A, et al. 2019.** Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**.
- Arkhipova I, Meselson M. 2005.** Deleterious transposable elements and the extinction of asexuals. *BioEssays* **27**: 76–85.
- Ashapkin V V., Kutueva LI, Vanyushin BF. 2002.** The gene for domains rearranged methyltransferase (DRM2) in *Arabidopsis thaliana* plants is methylated at both cytosine and adenine residues. *FEBS Letters* **532**: 367–372.
- Ausin I, Feng S, Yu C, Liu W, Yu H, Jacobsen EL, Zhai J. 2016.** DNA methylome of the 20-gigabase Norway spruce genome. *Proc Natl Acad Sci USA*: 1–8.
- Avramidou E V., Ganopoulos I V., Doulis AG, Tsaftaris AS, Aravanopoulos FA. 2015.** Beyond population genetics: natural epigenetic variation in wild cherry (*Prunus avium*). *Tree Genetics & Genomes* **11**: 95.
- Avramova Z. 2015.** Transcriptional ‘memory’ of a stress: Transient chromatin and memory (epigenetic) marks at stress-response genes. *Plant Journal* **83**: 149–159.
- Bahar E, Korkutal I, Carbonneau A, Akcay G. 2010.** Using magnetic resonance imaging technique (MRI) to investigate graft connection and its relation to reddening discoloration in grape leaves. *Journal of Food, Agriculture and Environment* **8**: 293–297.
- Ballouz S, Dobin A, Gillis JA. 2019.** Is it time to change the reference genome? *Genome Biology* **20**: 1–9.
- Ban Y, Honda C, Hatsuyama Y, Igarashi M, Bessho H, Moriguchi T. 2007.** Isolation and Functional Analysis of a MYB Transcription Factor Gene that is a Key Regulator for the Development of Red Coloration in Apple Skin. *Plant and Cell Physiology* **48**: 958–970.
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007.** CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)* **315**: 1709–12.

- Bartee L, Bender J. 2001.** Two Arabidopsis methylation-deficiency mutations confer only partial effects on a methylated endogenous gene family. *Nucleic Acids Research* **29**: 2127–2134.
- Bartel DP. 2004.** MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116**: 281–297.
- Basheer-Salimia R. 2007.** Juvenility, Maturation, and Rejuvenation in Woody Plants. *Horticultural Reviews*: 109–155.
- Baurens FC, Nicolleau J, Legavre T, Verdeil JL, Monteuis O. 2004.** Genomic DNA methylation of juvenile and mature *Acacia mangium* micropropagated in vitro with reference to leaf morphology as a phase change marker. *Tree Physiology* **24**: 401–407.
- Baylin SB. 2005.** DNA methylation and gene silencing in cancer. *Nature Clinical Practice Oncology* **2**: S4–S11.
- Bennetzen JL. 2005.** Transposable elements, gene creation and genome rearrangement in flowering plants. *Current Opinion in Genetics and Development* **15**: 621–627.
- Berger SL. 2007.** The complex language of chromatin regulation during transcription. *Nature* **447**: 407–412.
- Berger MMJ, Gallusci P, Teyssier E. 2018.** *Roles of Epigenetic Mechanisms in Grafting and Possible Applications*. Elsevier Ltd.
- Berry S, Dean C. 2015.** Environmental perception and epigenetic memory: Mechanistic insight through FLC. *Plant Journal* **83**: 133–148.
- Bertani G. 1951.** Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of bacteriology* **62**: 293–300.
- Bharat SS, Li S, Li J, Yan L, Xia L. 2019.** Base editing in plants: Current status and challenges. *Crop Journal*.
- Bird AP. 1984.** Gene expression: DNA methylation - How important in gene control? *Nature* **307**: 503–504.
- Black DL. 2003.** Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annual Review of Biochemistry* **72**: 291–336.
- Bocchini M, Bartucca ML, Ciancaleoni S, Mimmo T, Cesco S, Pii Y, Albertini E, Del Buono D. 2015.** Iron deficiency in barley plants: Phytosiderophore release, iron translocation, and DNA methylation. *Frontiers in Plant Science* **6**: 1–12.
- Bond BJ. 2000.** Age-related changes in photosynthesis of woody plants. *Trends in Plant Science* **5**: 349–353.
- Boyko A, Blevins T, Yao Y, Golubov A, Bilichak A, Ilnytskyy Y, Hollander J, Meins F, Kovalchuk I. 2010.** Transgenerational Adaptation of Arabidopsis to Stress Requires DNA Methylation and the Function of Dicer-Like Proteins (S-H Shiu, Ed.). *PLoS ONE* **5**: e9514.
- Bureau TE, Wessler SR. 1994.** Stowaway: A new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell* **6**: 907–916.
- Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, Gardner R, Berger F,**

- Feijó JA, Becker JD, et al. 2012.** Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA.
- Canadian Food Inspection Agency. 2020.** The Biology of *Malus domestica* Borkh. (Apple).
- Cao X, Jacobsen SE. 2002.** Role of the *Arabidopsis* DRM Methyltransferases in De Novo DNA Methylation and Gene Silencing transferases, the DRMs show a rearranged structure within their methyltransferase catalytic domains such that motifs VI through X are N-terminal to motifs I through.
- Carels N, Bernardi G. 1999.** Two classes of genes in plants. *Genetics* **154**: 1819–1825.
- Carisse O, Khanizadeh S. 2006.** Relative resistance of newly released apple rootstocks to *Phytophthora cactorum*. *Canadian Journal of Plant Science* **86**: 199–204.
- Casadesús J, Low DA. 2013.** Programmed heterogeneity: epigenetic mechanisms in bacteria. *The Journal of biological chemistry* **288**: 13929–35.
- Cedar H. 1988.** DNA methylation and gene activity. *Cell* **53**: 3–4.
- Celton JM, Dheilly E, Guillou MC, Simonneau F, Juchaux M, Costes E, Laurens F, Renou JP. 2014.** Additional amphivasal bundles in pedicel pith exacerbate central fruit dominance and induce self-thinning of lateral fruitlets in apple. *Plant Physiology* **164**: 1930–1951.
- Chan SW-L, Henderson IR, Jacobsen SE. 2005.** Gardening The Genome: DNA Methylation in *Arabidopsis thaliana*. *Nature Reviews Genetics* **6**: 590–590.
- Charrier A, Vergne E, Dousset N, Richer A, Petiteau A, Chevreau E. 2019.** Efficient targeted mutagenesis in apple and first time edition of pear using the CRISPR-Cas9 system. *Frontiers in Plant Science* **10**.
- Chen LT, Luo M, Wang YY, Wu K. 2010.** Involvement of *Arabidopsis* histone deacetylase HDA6 in ABA and salt stress response. *Journal of Experimental Botany* **61**: 3345–3353.
- Chen J, Tian L. 2007.** Roles of dynamic and reversible histone acetylation in plant. *Biochim Biophys Acta* **1769**: 295–307.
- Chevreau E, Dousset N, Joffrion C, Richer A, Charrier A, Vergne E. 2019.** Agroinfiltration is a key factor to improve the efficiency of apple and pear transformation. *Scientia Horticulturae* **251**: 150–154.
- Chien JC, Sussex IM. 1996.** Differential Regulation of Trichome Formation on the Adaxial and Abaxial Leaf Surfaces by Gibberellins and Photoperiod in *Arabidopsis thaliana* (L.) Heynh. *Plant Physiol* **111**: 1321–1328.
- Chopra S, Brendel V, Zhang J, Axtell JD, Peterson T. 1999.** Molecular characterization of a mutable pigmentation phenotype and isolation of the first active transposable element from *Sorghum bicolor*. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 15330–15335.
- Coart E, Van Glabeke S, De Loose M, Larsen AS, Roldán-Ruiz I. 2006.** Chloroplast diversity in the genus *Malus*: New insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). *Molecular Ecology* **15**: 2171–2182.

- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008a.** Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**: 215–219.
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008b.** Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**: 215–219.
- Company Stark Brothers. 1913.** Golden Delicious Apple.
- Cookson SJ, Clemente Moreno MJ, Hevin C, Nyamba Mendome LZ, Delrot S, Trossat-Magnin C, Ollat N. 2013.** Graft union formation in grapevine induces transcriptional changes related to cell wall modification, wounding, hormone signalling, and secondary metabolism. *Journal of Experimental Botany* **64**: 2997–3008.
- Cookson SJ, Ollat N. 2013.** Grafting with rootstocks induces extensive transcriptional re-programming in the shoot apical meristem of grapevine. *BMC Plant Biology* **13**: 147.
- Cornille A, Giraud T, Smulders MJM, Roldán-Ruiz I, Gladieux P. 2014.** The domestication and evolutionary ecology of apples. *Trends in Genetics* **30**: 57–65.
- Cornille A, Gladieux P, Smulders MJM, Roldán-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olova M, Feugey L, et al. 2012.** New Insight into the History of Domesticated Apple: Secondary Contribution of the European Wild Apple to the Genome of Cultivated Varieties (R Mauricio, Ed.). *PLoS Genetics* **8**: e1002703.
- Corpet F. 1988.** Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res* **16**: 10881–10890.
- Corredoira E, Cano V, Bárány I, Solís M-T, Rodríguez H, Vieitez A-M, Risueño MC, Testillano PS. 2017.** Initiation of leaf somatic embryogenesis involves high pectin esterification, auxin accumulation and DNA demethylation in *Quercus alba*. *Journal of Plant Physiology* **213**: 42–54.
- Crick F. 1990.** chapter 8: The genetic code. In: *What Mad Pursuit: A Personal View of Scientific Discovery*. 89–101.
- Criqui MC, Genschik P. 2002.** Mitosis in plants: How far we have come at the molecular level? *Current Opinion in Plant Biology* **5**: 487–493.
- Di Croce L, Raker VA, Corsaro M, Fazi F, Fanelli M, Faretta M, Fuks F, Lo Coco F, Kouzarides T, Nervi C, et al. 2002.** Methyltransferase Recruitment and DNA Hypermethylation of Target Promoters by an Oncogenic Transcription Factor. *Science* **295**: 1079–1082.
- Daccord N. 2018.** Analyse bioinformatique du génome et de l'épigénome du pommier.
- Daccord N, Celton J-M, Linsmith G, Becker C, Choisne N, Schijlen E, van de Geest H, Bianco L, Micheletti D, Velasco R, et al. 2017.** High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics*.
- Dalin P, Ågren J, Björkman C, Huttunen P, Kärkkäinen K. 2008.** Leaf trichome formation and plant resistance to herbivory. In: *Induced Plant Resistance to Herbivory*. Dordrecht: Springer Netherlands, 89–105.

- Darneli JE. 1978.** Implications of RNA * RNA Splicing in Evolution of Eukaryotic Cells. *Science* **202**.
- Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. 2011.** CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602–607.
- Desvoyes B, Fernández-Marcos M, Sequeira-Mendes J, Otero S, Vergara Z, Gutierrez C. 2014.** Looking at plant cell cycle from the chromatin window. *Frontiers in Plant Science* **5**: 1–12.
- Do PT, Nguyen CX, Bui HT, Tran LTN, Stacey G, Gillman JD, Zhang ZJ, Stacey MG. 2019.** Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous GmFAD2-1A and GmFAD2-1B genes to yield a high oleic, low linoleic and α -linolenic acid phenotype in soybean. *BMC Plant Biology* **19**: 1–14.
- Dobosy JR, Selker EU. 2001.** Emerging connections between DNA methylation and histone acetylation. *Cellular and molecular life sciences : CMLS* **58**: 721–7.
- Doudna JA, Charpentier E. 2014.** Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)* **346**: 1258096-1-1258096–9.
- Downen RH, Pelizzola M, Schmitz RJ, Lister R, Downen JM, Nery JR, Dixon JE, Ecker JR. 2012a.** Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E2183-91.
- Downen RH, Pelizzola M, Schmitz RJ, Lister R, Downen JM, Nery JR, Dixon JE, Ecker JR. 2012b.** Widespread dynamic DNA methylation in response to biotic stress. *Proceedings of the National Academy of Sciences of the United States of America* **109**.
- Durand S, Bouché N, Perez Strand E, Loudet O, Camilleri C. 2012.** Rapid establishment of genetic incompatibility through natural epigenetic variation. *Current Biology* **22**: 326–331.
- Eid A, Alshareef S, Mahfouz MM. 2018.** CRISPR base editors: Genome editing without double-stranded breaks. *Biochemical Journal* **475**: 1955–1964.
- Engler C, Kandzia R, Marillonnet S. 2008.** A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**: e3647.
- Espley R V., Hellens RP, Putterill J, Stevenson DE, Kutty-Amma S, Allan AC. 2007.** Red colouration in apple fruit is due to the activity of the MYB transcription factor, MdMYB10. *The Plant Journal* **49**: 414–427.
- Evers CA, Starr C, Starr L. 2013.** *Biology: Today and Tomorrow With Physiology, Forth Edition.*
- Faize M, Malnoy M, Dupuis F, Chevalier M, Parisi L, Chevreau E. 2003.** Chitinases of {Trichoderma} atroviride induce scab resistance and some metabolic changes in two cultivars of apple. *Phytopathology* **93**: 1496–1504.
- Fallahi E, Colt WM, Fallahi B, Chun IJ. 2002.** The importance of apple rootstocks on tree growth, yield, fruit quality, leaf nutrition, and photosynthesis with an emphasis on ‘FUJI’. *HortTechnology* **12**: 38–44.
- Fan S, Wang J, Lei C, Gao C, Yang Y, Li Y, An N, Zhang D, Han M. 2018.** Identification and characterization of histone modification gene family reveal their critical responses to flower

induction in apple. *BMC Plant Biology* **18**: 1–21.

Fauser F, Schiml S, Puchta H. 2014. Both CRISPR/Cas-based nucleases and nickases can be used efficiently for genome engineering in *Arabidopsis thaliana*. *The Plant Journal* **79**: 348–359.

Felton A, Nilsson U, Sonesson J, Felton AM, Roberge J-M, Ranius T, Ahlström M, Bergh J, Björkman C, Boberg J, et al. 2016. Replacing monocultures with mixed-species stands: Ecosystem service implications of two production forest alternatives in Sweden. *Ambio* **45**: 124–139.

Feng Y, Zhang X, Wu T, Xu X, Han Z, Wang Y. 2016. Methylation effect on IPT5b gene expression determines cytokinin biosynthesis in apple rootstock. *Biochemical and Biophysical Research Communications*: 1–6.

Feschotte C, Jiang N, Wessler SR. 2002. Plant transposable elements: Where genetics meets genomics. *Nature Reviews Genetics* **3**: 329–341.

Finnegan EJ, Genger RK, Peacock WJ, Dennis ES. 1998. DNA Methylation in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **49**: 223–247.

Finnegan EJ, Peacock WJ, Dennis ES. 1996. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 8449–8454.

van der Fits L, Deakin EA, Hoge JH, Memelink J. 2000. The ternary transformation system: constitutive virG on a compatible plasmid dramatically increases *Agrobacterium*-mediated plant transformation. *Plant molecular biology* **43**: 495–502.

Flachowsky H, Tränkner C, Szankowski I, Waidmann S, Hanke MV, Treutter D, Fischer TC. 2012. RNA-mediated gene silencing signals are not graft transmissible from the rootstock to the scion in greenhouse-grown apple plants *Malus* sp. *International Journal of Molecular Sciences* **13**: 9992–10009.

Fortes AM, Gallusci P. 2017. Plant Stress Responses and Phenotypic Plasticity in the Epigenomics Era: Perspectives on the Grapevine Scenario, a Model for Perennial Crop Plants. *Frontiers in Plant Science* **08**: 1–9.

Forum Reasearch Gate. 2013. No Title. *Research Gate*.

Foucher F, Morin J, Courtiade J, Cadioux S, Ellis N, Banfield MJ, Rameau C. 2003. Determinate and Late Flowering Are Two Terminal Flower1/Centroradialis Homologs That Control Two Distinct Phases of Flowering Initiation and Development in Pea. *Plant Cell* **15**: 2742–2754.

Fraga MF, Rodriguez R, Canal MJ. 2002. Genomic DNA methylation-demethylation during aging and reinvigoration of *Pinus radiata*. *Tree Physiology* **22**: 813–816.

Fuentes I, Stegemann S, Golczyk H, Karcher D, Bock R. 2014. Horizontal genome transfer as an asexual path to the formation of new species. *Nature*: 232–235.

Fulton TM, Chunwongse J, Tanksley SD. 1995. Microprep protocol for extraction of DNA from tomato and other herbaceous plants. *Plant Molecular Biology Reporter* **13**: 207–209.

Garcia-Diaz M, Bebenek K. 2007. Multiple functions of DNA polymerases. *Critical Reviews in Plant Sciences* **26**: 105–122.

- Gascioli V, Mallory AC, Bartel DP, Vaucheret H. 2005.** Partially Redundant Functions of Arabidopsis DICER-like Enzymes and a Role for DCL4 in Producing trans-Acting siRNAs. *Current Biology* **15**: 1494–1500.
- Gautier AT, Chambaud C, Brocard L, Ollat N, Gambetta GA, Delrot S, Cookson SJ. 2019.** Merging genotypes: Graft union formation and scion-rootstock interactions. *Journal of Experimental Botany* **70**: 805–815.
- Gehring M, Bubb KL, Henikoff S. 2009.** Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* **324**: 1447–1451.
- Goble AM, Fan H, Sali A, Raushel FM. 2011.** Discovery of a cytokinin deaminase. *ACS chemical biology* **6**: 1036–40.
- Goll MG, Bestor TH. 2005.** Eukaryotic Cytosine Methyltransferases. *Annual Review of Biochemistry* **74**: 481–514.
- González RM, Ricardi MM, Iusem ND. 2013.** Epigenetic marks in an adaptive water stress-responsive gene in tomato roots under normal and drought conditions. *Epigenetics* **8**: 864–872.
- van der Graaf A, Wardenaar R, Neumann DA, Taudt A, Shaw RG, Jansen RC, Schmitz RJ, Colomé-Tatché M, Johannes F. 2015.** Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proceedings of the National Academy of Sciences* **112**: 6676–6681.
- Greenwood MS. 1995.** Juvenility and maturation in conifers : current concepts. : 433–438.
- Gregory PJ, Atkinson CJ, Bengough AG, Else MA, Fernández-Fernández F, Harrison RJ, Schmidt S. 2013.** Contributions of roots and rootstocks to sustainable, intensified crop production. *Journal of Experimental Botany* **64**: 1209–1222.
- Griffiths A, Miller J, Suzuki D. 2000.** Bacterial conjugation. In: Freeman W, ed. An introduction to genetic analysis. New York.
- Gross BL, Henk AD, Richards CM, Fazio G, Volk GM. 2014.** Genetic diversity in *Malus × domestica* (Rosaceae) through time in response to domestication. *American journal of botany* **101**: 1770–9.
- Gully K, Celton J-M, Degrave A, Pelletier S, Brisset M-N, Bucher E. 2019a.** Biotic Stress-Induced Priming and De-Priming of Transcriptional Memory in Arabidopsis and Apple. *Epigenomes* **3**: 3.
- Gully K, Pelletier S, Guillou MC, Ferrand M, Aligon S, Pokotylo I, Perrin A, Vergne E, Fagard M, Ruelland E, et al. 2019b.** The SCOOP12 peptide regulates defense response and root elongation in Arabidopsis thaliana. *Journal of Experimental Botany* **70**: 1349–1365.
- Guy CP, Majerník AI, Chong JPJ, Bolt EL. 2004.** A novel nuclease-ATPase (Nar71) from archaea is part of a proposed thermophilic DNA repair system. *Nucleic Acids Research* **32**: 6176–6186.
- Haag JR, Ream TS, Marasco M, Nicora CD, Norbeck AD, Pasa-Tolic L, Pikaard CS. 2012.** In Vitro Transcription Activities of Pol IV, Pol V, and RDR2 Reveal Coupling of Pol IV and RDR2 for dsRNA Synthesis in Plant RNA Silencing. *Molecular Cell* **48**: 811–818.
- Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud JB, Schneider-**

- Maunoury S, Shkumatava A, Teboul L, Kent J, et al. 2016.** Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology* **17**: 1–12.
- Hagmann J, Becker C, Müller J, Stegle O, Meyer RC, Wang G, Schneeberger K, Fitz J, Altmann T, Bergelson J, et al. 2015.** Century-scale Methylome Stability in a Recently Diverged *Arabidopsis thaliana* Lineage. *PLoS Genetics* **11**: 1004920.
- Hanke M-V, Flachowsky H, Peil A, Hättasch C. 2007.** No flower no fruit—genetic potentials to trigger flowering in fruit trees. *Genes Genomes Genomics* **1**: 1–20.
- Haraguchi S, Nakagawara A. 2009.** A simple PCR method for rapid genotype analysis of the TH-MYCN transgenic mouse. *PLoS ONE* **4**: 3–6.
- Harris SA, Julian P, Juniper BE. 2002.** Genetics and the origin of apples. **18**: 426–430.
- Hartley JL, Temple GF, Brasch MA. 2000.** DNA Cloning Using In Vitro Site-Specific Recombination. : 1788–1795.
- Harvey L, Berk A, Kaiser CA, Krieger M, Scott MP, Bretscher A, Matsudaira P. 2007.** chapter 3. Protein Structure and Function. In: *Molecular Cell Biology*. W. H. Freeman; 6th edition.
- Hasbún R, Valledor L, Berdasco M, Santamaría E, Cañal MJ, Rodríguez R, Ríos D, Sánchez M. 2005.** In vitro proliferation and genome dna methylation in adult chestnuts. *Acta Horticulturae* **693**: 333–340.
- Hauben M, Haesendonckx B, Standaert E, Van Der Kelen K, Azmi A, Akpo H, Van Breusegem F, Guisez Y, Bots M, Lambert B, et al. 2009.** Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 20109–20114.
- Hauser MT, Aufsatz W, Jonak C, Luschnig C. 2011.** Transgenerational epigenetic inheritance in plants. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1809**: 459–468.
- He X-J, Chen T, Zhu J-K. 2011.** Regulation and function of DNA methylation in plants and animals. *Cell Research* | **21**: 442–465.
- He P, Cheng L, Li H, Wang H, Li L. 2017.** A comparative analysis of DNA methylation in diploid and tetraploid apple (*Malus × domestica* borkh.). *Czech Journal of Genetics and Plant Breeding* **53**: 63–68.
- Henderson IR, Jacobsen SE. 2007.** Epigenetic inheritance in plants. *Nature*.
- Herman JJ, Sultan SE. 2011.** Adaptive transgenerational plasticity in plants: case studies, mechanisms, and implications for natural populations. *Frontiers in plant science* **2**: 102.
- Herrera CM, Bazaga P. 2016.** Genetic and epigenetic divergence between disturbed and undisturbed subpopulations of a Mediterranean shrub: A 20-year field experiment. *Ecology and Evolution* **6**: 3832–3847.
- Hiyoshi A, Miyahara K, Kato C, Ohshima Y. 2011.** Does a DNA-less cellular organism exist on Earth? *Genes to Cells* **16**: 1146–1158.

- Hollister JD, Gaut BS. 2009.** Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research* **19**: 1419–1428.
- Honda C, Kotoda N, Wada M, Kondo S, Kobayashi S, Soejima J, Zhang Z, Tsuda T, Moriguchi T. 2002.** Anthocyanin biosynthetic genes are coordinately expressed during red coloration in apple skin. *Plant Physiology and Biochemistry* **40**: 955–962.
- Hood EE, Gelvin SB, Melchers LS, Hoekema A. 1993.** New *Agrobacterium* helper plasmids for gene transfer to plants. *Transgenic Research* **2**: 208–218.
- House M, Lukens L. 2019.** Chapter 3 The Role of germinally Inherited Epialleles in Plant Breeding: An Update. In: *Epigenetics in Plants of Agronomic Importance: Fundamentals and Applications*. 115–125.
- Hsieh T-F, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D. 2009.** Genome-Wide Demethylation of *Arabidopsis* Endosperm. *Science* **324**: 1451–1454.
- Hutchison KW, Sherman CD, Weber J, Smith SS, Singer PB, Greenwood MS. 1990.** Maturation in Larch: II. Effects of Age on Photosynthesis and Gene Expression in Developing Foliage. *Plant Physiology* **94**: 1308–1315.
- Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang TH, Lan T, Welch AJ, Juárez MJA, Simpson J, et al. 2013.** Architecture and evolution of a minute plant genome. *Nature* **498**: 94–98.
- Ibarra CA, Feng X, Schoft VK, Hsieh TF, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T, et al. 2012.** Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337**: 1360–1364.
- Inamdar NM, Ehrlich KC, Ehrlich M. 1991.** CpG methylation inhibits binding of several sequence-specific DNA-binding proteins from pea, wheat, soybean and cauliflower. *Plant Molecular Biology* **17**: 111–123.
- Iwasaki M, Paszkowski J, Freeling M. 2014.** Epigenetic memory in plants. *The EMBO journal* **33**: 1987–98.
- Jackson JP, Lindroth AM, Cao X, Jacobsen SE. 2002.** Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* **416**: 556–560.
- Janick J, Cummins JN, Brown SK, Hemmat M. 1996.** Chapter I Apples. In: Janick J, Moore JN, eds. *Tree and Tropical Fruits*. John Wiley & Sons, Inc, 239–248.
- Jensen PJ, Halbrecht N, Fazio G, Makalowska I, Altman N, Praul C, Maximova SN, Ngugi HK, Crassweller RM, Travis JW, et al. 2012.** Rootstock-regulated gene expression patterns associated with fire blight resistance in apple. *BMC Genomics* **13**.
- Jensen PJ, Makalowska I, Altman N, Fazio G, Praul C, Maximova SN, Crassweller RM, Travis JW, McNellis TW. 2010.** Rootstock-regulated gene expression patterns in apple tree scions. *Tree Genetics and Genomes* **6**: 57–72.
- Joyce SM, Cassells AC, Jain SM. 2003.** Stress and quality in in vitro culture. *Plant Cell, Tissue and Organ Culture* **74**: 103–121.

- Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F. 2012.** DNA Methylation Dynamics during Sexual Reproduction in *Arabidopsis thaliana*. *Current Biology* **22**: 1825–1830.
- Kakutani T, Jeddeloh JA, Flowers SK, Munakata K, Richards EJ. 1996.** Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 12406–12411.
- Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ. 2002.** *Arabidopsis* MET1 cytosine methyltransferase mutants. *Genetics* **163**: 1109–1122.
- Karnatz A. 1963.** Der Blattrand als Selektionsmerkmal bei Apfelsamlingen. *Der Züchter* **7**.
- Kawakatsu T, Huang S, Shan C, Carol, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, et al. 2016.** Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cell* **166**: 492–506.
- Kawakatsu T, Nery JR, Castanon R, Ecker JR. 2017.** Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biology* **18**: 171.
- Kawashima T, Berger F. 2014.** Epigenetic reprogramming in plant sexual reproduction. *Nature Reviews Genetics* **15**: 613–624.
- Keyte AL, Percifield R, Liu B, Wendel JF. 2006.** Intraspecific DNA methylation polymorphism in cotton (*Gossypium hirsutum* L.). *Journal of Heredity* **97**: 444–450.
- Kidwell MG. 2002.** Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**: 49–63.
- Kim J, Kim JH, Richards EJ, Chung KM, Woo HR. 2014.** *Arabidopsis* VIM Proteins Regulate Epigenetic Silencing by Modulating DNA Methylation and Histone Modification in Cooperation with MET1. *Molecular Plant* **7**: 1470–1485.
- Kim M, Ohr H, Lee JW, Hyun Y, Fischer RL, Choi Y. 2008.** Temporal and spatial downregulation of *Arabidopsis* MET1 activity results in global DNA hypomethylation and developmental defects. *Molecules and cells* **26**: 611–5.
- Kinoshita T. 2004.** One-Way Control of FWA Imprinting in *Arabidopsis* Endosperm by DNA Methylation. *Science* **303**: 521–523.
- Kinoshita T, Seki M. 2014.** Epigenetic memory for stress response and adaptation in plants. *Plant and Cell Physiology* **55**: 1859–1863.
- Kotoda N, Wada M, Masuda T, Soejima J. 2003.** The break-through in the reduction of juvenile phase in apple using transgenic approaches. *Acta Horticulturae* **625**: 337–343.
- Kumari A, Kumar J, Kumar A, Chaudhury A, Singh SP. 2015.** Grafting triggers differential responses between scion and rootstock. *PLoS ONE* **10**: 1–19.
- De La Chaux N, Tsuchimatsu T, Shimizu KK, Wagner A. 2012.** The predominantly selfing plant *Arabidopsis thaliana* experienced a recent reduction in transposable element abundance compared to its outcrossing relative *Arabidopsis lyrata*. *Mobile DNA* **3**: 1–18.
- Laible M, Boonrod K. 2009.** Homemade site directed mutagenesis of whole plasmids. *Journal of*

visualized experiments : JoVE.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics* **11**: 204–220.

Le TN, Schumann U, Smith NA, Tiwari S, Khang Au PC, Zhu QH, Taylor JM, Kazan K, Llewellyn DJ, Zhang R, et al. 2014. DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in Arabidopsis. *Genome Biology* **15**: 1–18.

Lee R, Feinbaum R, Ambros V. 1993. The *C. elegans* Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*. *Cell* **75**: 843–854.

Lei Y, Zhang X, Su J, Jeong M, Gundry MC, Huang Y-H, Zhou Y, Li W, Goodell MA. 2017. Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. *Nature Communications* **8**: 16026.

Lespinasse Y, Bouvier L, Djulbic M, Chevreau E. 1999. Haploidy in apple and pear. *Acta Horticulturae* **484**: 49–54.

Li H, Geng M, Liu Q, Jin C, Zhang Q, Chen C, Song W, Wang C. 2014. Characteristics of cytosine methylation status and methyltransferase genes in the early development stage of cauliflower (*Brassica oleracea* L. var. botrytis). *Plant Cell, Tissue and Organ Culture* **117**: 187–199.

Li X, Harris CJ, Zhong Z, Chen W, Liu R, Jia B, Wang Z, Li S, Jacobsen SE, Du J. 2018. Mechanistic insights into plant SUVH family H3K9 methyltransferases and their binding to context-biased non-CG DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America* **115**: E8793–E8802.

Li X, Kui L, Zhang J, Xie Y, Wang L, Yan Y, Wang N, Xu J, Li C, Wang W, et al. 2016. Improved hybrid de novo genome assembly of domesticated apple (*Malus x domestica*). *GigaScience* **5**.

Li W, Liu H, Cheng ZJ, Su YH, Han HN, Zhang Y, Zhang XS. 2011. DNA methylation and histone modifications regulate de novo shoot regeneration in Arabidopsis by modulating WUSCHEL expression and auxin signaling. *PLoS genetics* **7**: e1002243.

Lindroth AM. 2001. Requirement of CHROMOMETHYLASE3 for Maintenance of CpXpG Methylation. *Science* **292**: 2077–2080.

Lister R, Ecker JR. 2009. Finding the fifth base: Genome-wide sequencing of cytosine methylation. *Genome Research* **19**: 959–966.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly Integrated Single-Base Resolution Maps of the Epigenome in Arabidopsis. *Cell* **133**: 523–536.

Liu Q, Feng Y, Zhu Z. 2009. Dicer-like (DCL) proteins in plants. *Functional and Integrative Genomics* **9**: 277–286.

Liu Y, Ma S, Chang J, Zhang T, Chen X, Liang Y, Xia Q. 2019. Programmable targeted epigenetic editing using CRISPR system in *Bombyx mori*. *Insect Biochemistry and Molecular Biology* **110**: 105–111.

- Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R. 2016.** Editing DNA Methylation in the Mammalian Genome. *Cell* **167**: 233-247.e17.
- Liu H, Zhang H, Dong YX, Hao YJ, Zhang XS. 2018.** DNA METHYLTRANSFERASE1-mediated shoot regeneration is regulated by cytokinin-induced cell cycle in Arabidopsis. *New Phytologist* **217**: 219–232.
- Lohse M, Nagel A, Herter T, May P, Schroda M, Zrenner R, Tohge T, Fernie AR, Stitt M, Usadel B. 2014.** Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell & Environment* **37**: 1250–1258.
- Loidl P. 2004.** A plant dialect of the histone language. *Trends in Plant Science* **9**: 84–90.
- Lowder LG, Zhang D, Baltus NJ, Paul JW, Tang X, Zheng X, Voytas DF, Hsieh T-F, Zhang Y, Qi Y. 2015.** A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome Editing and Transcriptional Regulation. *Plant physiology* **169**: 971–85.
- Luo M, Liu X, Singh P, Cui Y, Zimmerli L, Wu K. 2012.** Chromatin modifications and remodeling in plant abiotic stress responses. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* **1819**: 129–136.
- Ma C, Jing C, Chang B, Yan J, Liang B, Liu L, Yang Y, Zhao Z. 2018.** The effect of promoter methylation on MdMYB1 expression determines the level of anthocyanin accumulation in skins of two non-red apple cultivars. *BMC Plant Biology* **18**: 1–14.
- Maden BEH, Hughes JMX. 1997.** Eukaryotic ribosomal RNA: The recent excitement in the nucleotide modification problem. *Chromosoma* **105**: 391–400.
- Makarova KS. 2002.** A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Research* **30**: 482–496.
- Makarova KS, Grishin N V, Shabalina SA, Wolf YI, Koonin E V. 2006.** A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct* **1**: 7.
- Malagnac F, Bartee L, Bender J. 2002.** An Arabidopsis SET domain protein required for maintenance but not establishment of DNA methylation. *The EMBO journal* **21**: 6842–52.
- Malnoy M, Viola R, Jung M-H, Koo O-J, Kim S, Kim J-S, Velasco R, Nagamangala Kanchiswamy C. 2016.** DNA-Free Genetically Edited Grapevine and Apple Protoplast Using CRISPR/Cas9 Ribonucleoproteins. *Frontiers in Plant Science* **7**: 1904.
- Manova V, Gruszka D. 2015.** DNA damage and repair in plants – From models to crops. *Frontiers in Plant Science* **6**: 1–26.
- Martin C, Paz-Ares J. 1997.** MYB transcription factors in plants. *Trends in Genetics* **13**: 67–73.
- Mateo-Bonmatí E, Casanova-Sáez R, Ljung K. 2019.** Epigenetic Regulation of Auxin Homeostasis. *Biomolecules* **9**: 623.
- Mcclelland M, Nelson M, Raschke E. 1994.** Effect of site-specific modification on restriction endonucleases and DNA modification methyltransferases. *Nucleic Acids Research* **22**: 3640–3659.

- McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, Kramer A, Martens A, Edwardsand JR, Challen GA. 2016.** Reprogrammable CRISPR/Cas9-based system for inducing sitespecific DNA methylation. *Biology Open* **5**: 866–874.
- McKey D, Elias M, Pujol ME, Duputié A. 2010.** The evolutionary ecology of clonally propagated domesticated plants. *New Phytologist* **186**: 318–332.
- Melnyk CW. 2017a.** Plant grafting: insights into tissue regeneration. *Regeneration* **4**: 3–14.
- Melnyk CW. 2017b.** Connecting the plant vasculature to friend or foe. *New Phytologist* **213**: 1611–1617.
- Melnyk C, Gabel A, Hardcastle T, Robinson S, Miyashima S, Grosse I, Meyerowitz E. 2018.** Transcriptome dynamics at Arabidopsis graft junctions reveal an intertissue recognition mechanism that activates vascular regeneration. *Proceedings of the National Academy of Sciences*.
- Melnyk CW, Meyerowitz EM. 2015.** Plant grafting. *Current Biology* **25**: R183–R188.
- Mendel G. 1865.** Centenary of Mendel’s Paper. *Verhandlungen des naturforschenden Vereines*: 1–47.
- Mercier R, Mézard C, Jenczewski E, Macaisne N, Grelon M. 2015.** The Molecular Biology of Meiosis in Plants. *Annual Review of Plant Biology* **66**: 297–327.
- Michalowsky LA, Jones PA. 1989.** DNA methylation and differentiation. *Environmental Health Perspectives* **80**: 189–197.
- Milien M, Renault-Spilmont AS, Cookson SJ, Sarrazin A, Verdeil JL. 2012.** Visualization of the 3D structure of the graft union of grapevine using X-ray tomography. *Scientia Horticulturae* **144**: 130–140.
- Mimida N, Kotoda N, Ueda T, Igarashi M, Hatsuyama Y, Iwanami H, Moriya S, Abe K. 2009.** Four TFL1/CEN-like genes on distinct linkage groups show different expression patterns to regulate vegetative and reproductive development in apple (*Malus x domestica* Borkh.). *Plant and Cell Physiology* **50**: 394–412.
- Mirouze M, Paszkowski J. 2011.** Epigenetic contribution to stress adaptation in plants. *Current Opinion in Plant Biology* **14**: 267–274.
- Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, Cao J, Weigel D, Paszkowski J, Mathieu O. 2009.** Selective epigenetic control of retrotransposition in Arabidopsis. *Nature* **461**: 427–430.
- Moisan M. 2010.** Rat Genomics. *Business* **597**: 1–33.
- Monteuuis O, Doubeau S, Verdeil JL. 2008.** DNA methylation in different origin clonal offspring from a mature Sequoiadendron giganteum genotype. *Trees - Structure and Function* **22**: 779–784.
- Les Mordus de la Pomme. 2020.** Pommes à cidre - Les Mordus de la Pomme.
- Moreno MJC, Hevin C, Ollat N, Cookson SJ. 2014.** Developments at the graft interface in homo- and hetero-grafts gene expression and histological changes during the first month after grafting. *Plant Signaling and Behavior* **9**.

- Mosher RA, Lewsey MG, Shivaprasad P V. 2010.** RNA silencing in plants: Flash report! *Silence* **1**: 13.
- Mosher RA, Melnyk CW, Kelly KA, Dunn RM, Studholme DJ, Baulcombe DC. 2009.** Uniparental expression of PolIV-dependent siRNAs in developing endosperm of Arabidopsis. *Nature* **460**: 283–286.
- Mudge K, Janick J, Scofield S, Goldschmidt EE. 2009.** A History of Grafting. *Horticultural Reviews* **35**: 437–493.
- Murashige T, Skoog F. 1962.** A Revised Medium for Rapid Growth and Bio Assays with Tohaoco Tissue Cultures. *Physiologia Plantarum* **15**: 474–497.
- Nakajima I, Ban Y, Azuma A, Onoue N, Moriguchi T, Yamamoto T, Toki S, Endo M. 2017.** CRISPR/Cas9-mediated targeted mutagenesis in grape. *PLoS ONE* **12**: 1–16.
- Nanda AK, Melnyk CW. 2018.** The role of plant hormones during grafting. *Journal of Plant Research* **131**: 49–58.
- Newton AC, Fitt BDL, Atkins SD, Walters DR, Daniell TJ. 2010.** Pathogenesis, parasitism and mutualism in the trophic space of microbe–plant interactions. *Trends in Microbiology* **18**: 365–373.
- Niederhuth CE, Bewick AJ, Ji L, Alabady M, Kim K Do, Page JT, Li Q, Rohr NA, Rambani A, Burke JM, et al. 2016.** Widespread natural variation of DNA methylation within angiosperms. *Genome Biology*: 1–19.
- Niederhuth CE, Schmitz RJ. 2016.** Putting DNA methylation in context: From genomes to gene expression in plants. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*.
- Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014.** Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*. **156**: 935–949.
- Nishitani C, Hirai N, Komori S, Wada M, Okada K, Osakabe K, Yamamoto T, Osakabe Y. 2016.** Efficient Genome Editing in Apple Using a CRISPR/Cas9 system. *Scientific Reports* **6**: 31481.
- Noël E, Jarne P, Glémin S, MacKenzie A, Segard A, Sarda V, David P. 2017.** Experimental Evidence for the Negative Effects of Self-Fertilization on the Adaptive Potential of Populations. *Current Biology* **27**: 237–242.
- Onodera Y, Haag JR, Ream T, Nunes PC, Pontes O, Pikaard CS. 2005.** Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation- dependent heterochromatin formation. *Cell* **120**: 613–622.
- Orcheski B, Brown SK. 2012.** A Grower’s Guide to Self and Cross-Incompatibility in Apple. *New York Fruit Quarterly* **20**: 25–28.
- Orlova E V. 2009.** How viruses infect bacteria. *EMBO Journal* **28**: 797–798.
- Ossowski S, Schneeberger K, Lucas-Lledo JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. 2010.** The Rate and Molecular Spectrum of Spontaneous Mutations in Arabidopsis thaliana. *Science* **327**: 92–94.

- Papikian A, Liu W, Gallego-Bartolomé J, Jacobsen SE. 2019.** Site-specific manipulation of Arabidopsis loci using CRISPR-Cas9 SunTag systems. *Nature Communications* **10**: 729.
- Peng A, Chen S, Lei T, Xu L, He Y, Wu L, Yao L, Zou X. 2017.** Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus. *Plant Biotechnology Journal* **15**: 1509–1519.
- Perrin A, Daccord N, Roquis D, Celton J-M, Vergne E, Bucher E. 2020.** Divergent DNA Methylation Signatures of Juvenile Seedlings, Grafts and Adult Apple Trees. *Epigenomes* **4**.
- Peterson CL, Laniel M-A. 2004.** Histones and histone modifications. *Current Biology* **14**: R546–R551.
- Pikaard CS, Scheid OM. 2014.** Epigenetic regulation in plants. *Cold Spring Harbor Perspectives in Biology* **6**: 1–32.
- Pina A, Cookson S, Calatayud A, Trinchera A, Errea P. 2017.** Physiological and molecular mechanisms underlying graft compatibility. In: *Vetable grafting principles and practices*.
- Poires Association Nationale Pomme. 2019.** Production en France par variétés | La Pomme.
- Pompili V, Dalla Costa L, Piazza S, Pindo M, Malnoy M. 2020.** Reduced fire blight susceptibility in apple cultivars using a high-efficiency CRISPR/Cas9-FLP/FRT-based gene editing system. *Plant Biotechnology Journal* **18**: 845–858.
- Promis A, Allen RB. 2017.** Tree seedlings respond to both light and soil nutrients in a Patagonian evergreen-deciduous forest. *PloS one* **12**: e0188686.
- Puchta H, Fauser F. 2014.** Synthetic nucleases for genome engineering in plants: Prospects for a bright future. *Plant Journal* **78**: 727–741.
- Rakyan VK, Beck S. 2006.** Epigenetic variation and inheritance in mammals. *Current Opinion in Genetics and Development* **16**: 573–577.
- Ratel D, Ravanat J-L, Berger F, Wion D. 2006.** N6-methyladenine: the other methylated base of DNA. *BioEssays : news and reviews in molecular, cellular and developmental biology* **28**: 309–15.
- Richards AJ. 1997.** *Plant breeding systems*. Chapman & Hall.
- Richards CL, Bossdorf O, Verhoeven KJF. 2010.** Understanding natural epigenetic variation. *New Phytologist* **187**: 562–564.
- Ríos G, Leida C, Conejero A, Badenes ML. 2014.** Epigenetic regulation of bud dormancy events in perennial plants. *Frontiers in Plant Science* **5**: 1–6.
- Robinson T. 2008.** The evolution towards more competitive apple orchard systems in the USA. *Acta Horticulturae* **772**: 491–500.
- Robinson JP, Harris SA, Juniper BE. 2001.** Taxonomy of the genus *Malus* mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Systematics and Evolution* **226**: 35–58.
- Roda A, Nyrop J, English-Loeb G. 2003.** Leaf pubescence mediates the abundance of non-prey food and the density of the predatory mite *Typhlodromus pyri*. *Experimental and Applied Acarology* **29**: 193–211.

- Rodrigues JA, Zilberman D. 2015.** Evolution and function of genomic imprinting in plants. *Genes and Development* **29**: 2517–2531.
- Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL. 1996a.** Demethylation-induced developmental pleiotropy in Arabidopsis. *Science* **273**: 654–657.
- Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL. 1996b.** Demethylation-Induced Developmental Pleiotropy in Arabidopsis. *Science* **273**: 654–657.
- Rong YS, Golic KG. 2000.** Gene targeting by homologous recombination in Drosophila. *Science* **288**: 2013–2018.
- Rouphael Y, Kyriacou MC, Colla G. 2018.** Vegetable grafting: A toolbox for securing yield stability under multiple stress conditions. *Frontiers in Plant Science* **8**: 10–13.
- Russo VEA, Martienssen RA, Riggs AD. 1996.** *Epigenetic mechanisms of gene regulation*. Cold Spring Harbor Laboratory Press.
- Salmon A, Ainouche ML, Wendel JF. 2005.** Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). *Molecular Ecology* **14**: 1163–1175.
- Schalk C, Drevensek S, Kramdi A, Kassam M, Ahmed I, Cognat V, Graindorge S, Bergdoll M, Baumberger N, Heintz D, et al. 2016.** DNA DAMAGE BINDING PROTEIN 2 (DDB2) Shapes the DNA Methylation Landscape. *The Plant Cell* **2**: tpc.00474.2016.
- Schlathölder I, Jänsch M, Flachowsky H, Broggin GAL, Hanke MV, Patocchi A. 2018.** Generation of advanced fire blight-resistant apple (*Malus × domestica*) selections of the fifth generation within 7 years of applying the early flowering approach. *Planta* **247**: 1475–1488.
- Schlumbaum A, van Glabeke S, Roldan-Ruiz I. 2012.** Towards the onset of fruit tree growing north of the Alps: Ancient DNA from waterlogged apple (*Malus* sp.) seed fragments. *Annals of Anatomy* **194**: 157–162.
- Schmitz RJ, He Y, Valdés-López O, Khan SM, Joshi T, Urich MA, Nery JR, Diers B, Xu D, Stacey G, et al. 2013.** Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Research* **23**: 1663–1674.
- Schmitz RJ, Schultz MD, Lewsey MG, Malley RCO, Urich MA, Libiger O, Schork NJ, Ecker JR. 2012.** Transgenerational Epigenetic Instability Is a Source of Novel Methylation Variants. **334**: 369–373.
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009.** The B73 maize genome: Complexity, diversity, and dynamics. *Science* **326**: 1112–1115.
- Schöffl F, Raschke E, Nagao RT. 1984.** The DNA sequence analysis of soybean heat-shock genes and identification of possible regulatory promoter elements. *The EMBO Journal* **3**: 2491–2497.
- Seymour DK, Becker C. 2017.** The causes and consequences of DNA methylome variation in plants. *Current Opinion in Plant Biology* **36**: 56–63.
- Seymour DK, Koenig D, Hagmann J, Becker C, Weigel D. 2014.** Evolution of DNA Methylation Patterns in the Brassicaceae is Driven by Differences in Genome Organization. *PLoS Genetics* **10**.

- Shah SA, Erdmann S, Mojica FJM, Garrett RA. 2013.** Protospacer recognition motifs: Mixed identities and functional diversity. *RNA Biology* **10**: 891–899.
- Shannon S, Ry Meeks-Wagner O. 1991.** A Mutation in the Arabidopsis TFL1 Gene Affects Inflorescence Meristem Development. *The Plant Cell* **3**: 877–892.
- Sharp SJ, Schaack J, Cooley L, Burke DJ, Soil D. 1985.** Structure and transcription of eukaryotic tRNA gene. *Critical Reviews in Biochemistry and Molecular Biology* **19**: 107–144.
- Shu H, Wildhaber T, Siretskiy A, Gruissem W, Hennig L. 2012.** Distinct modes of DNA accessibility in plant chromatin. *Nature Communications* **3**.
- Simons RK. 1986.** Graft-Union Characteristics As Related To Dwarfing in Apple (*Malus Domestica* Borkh.). *Acta Horticulturae*: 57–66.
- Singh J, Fabrizio J, Desnoues E, Silva JP, Busch W, Khan A. 2019.** Root system traits impact early fire blight susceptibility in apple (*Malus × domestica*). *BMC Plant Biology* **19**: 1–14.
- Slaughter A, Daniel X, Flors V, Luna E, Hohn B, Mauch-Mani B. 2012.** Descendants of primed Arabidopsis plants exhibit resistance to biotic stress. *Plant Physiology* **158**: 835–843.
- Slotkin RK, Martienssen R. 2007.** Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* **8**: 272–285.
- Slotkin RK, Vaughn M, Borges F, Tanurdzić M, Becker JD, Feijó JA, Martienssen RA. 2009.** Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**: 461–72.
- Suzuki MM, Bird A. 2008.** DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews. Genetics* **9**: 465–76.
- TAIR. 2020.** The Arabidopsis Information Resource.
- Takos AM, Jaffé FW, Jacob SR, Bogs J, Robinson SP, Walker AR. 2006.** Light-induced expression of a MYB gene regulates anthocyanin biosynthesis in red apples. *Plant physiology* **142**: 1216–32.
- Talbert PB, Henikoff S. 2010.** Histone variants — ancient wrap artists of the epigenome. *Nature Reviews Molecular Cell Biology* **11**: 264–275.
- Thakore PI, D'Ippolito AM, Song L, Safi A, Shivakumar NK, Kabadi AM, Reddy TE, Crawford GE, Gersbach CA. 2015.** Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nature Methods* **12**: 1143–1149.
- Thiebaut F, Hemerly AS, Ferreira PCG. 2019.** A Role for Epigenetic Regulation in the Adaptation and Stress Responses of Non-model Plants. *Frontiers in Plant Science* **10**: 246.
- Thirulogachandar V, Pandey P, Vaishnavi CS, Reddy MK. 2011.** An affinity-based genome walking method to find transgene integration loci in transgenic genome. *Analytical Biochemistry* **416**: 196–201.
- Tjian R. 1996.** The biochemistry of transcription in eukaryotes : A paradigm for multisubunit regulatory complexes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **351**: 491–499.

- Valladares F, Niinemets Ü. 2008.** Shade Tolerance, a Key Plant Feature of Complex Nature and Consequences. *Annual Review of Ecology, Evolution, and Systematics* **39**: 237–257.
- Vanyushin BF, Ashapkin V V. 2006.** DNA methylation in plants. *Springer*: 1–152.
- Vaucheret H. 2008.** Plant ARGONAUTES. *Trends in Plant Science* **13**: 350–358.
- Velasco R, Zharkikh A, Affourtit J. 2010.** The genome of the domesticated apple (*Malus domestica* Borkh.). *Nature genetics* **42**: 833–839.
- Viggiano L, de Pinto MC. 2017.** Dynamic DNA Methylation Patterns in Stress Response. In: Springer, Cham, 281–302.
- Visser T. 1964.** *Juvenile phase and growth of apple and pear seedlings.*
- Vojta A, Dobrinic P, Tadic V, Bockor L, Korac P, Julg B, Klasic M, Zoldos V. 2016.** Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Research* **44**: 5615–5628.
- Walker J, Gao H, Zhang J, Aldridge B, Vickers M, Higgins JD, Feng X. 2018.** Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. *Nature Genetics* **50**: 130–137.
- Wang R, Farrona S, Vincent C, Joecker A, Schoof H, Turck F, Alonso-Blanco C, Coupland G, Albani MC. 2009.** PEP1 regulates perennial flowering in *Arabis alpina*. *Nature* **459**: 423–427.
- Wang X, Tu M, Wang D, Liu J, Li Y, Li Z, Wang Y, Wang X. 2018.** CRISPR/Cas9-mediated efficient targeted mutagenesis in grape in the first generation. *Plant Biotechnology Journal* **16**: 844–855.
- Wassenegger M, Heimes S, Riedel L, Sanger HL. 1994.** RNA-directed de novo methylation of genomic sequences in plants. *Cell* **76**: 567–576.
- Wei L, Gu L, Song X, Cui X, Lu Z, Zhou M, Wang L, Hu F, Zhai J, Meyers BC, et al. 2014.** Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 3877–3882.
- Wendel JF, Greilhuber J, Doležel J, Leitch IJ. 2012.** Plant genome diversity volume 1: Plant genomes, their residents, and their evolutionary dynamics. *Plant Genome Diversity Volume 1: Plant Genomes, their Residents, and their Evolutionary Dynamics*: 1–279.
- Wenzel S, Flachowsky H, Hanke MV. 2013.** The Fast-track breeding approach can be improved by heat-induced expression of the FLOWERING LOCUS T genes from poplar (*Populus trichocarpa*) in apple (*Malus × domestica* Borkh.). *Plant Cell, Tissue and Organ Culture* **115**: 127–137.
- Wessler SR. 1988.** Phenotypic diversity mediated by the maize transposable elements Ac and Spm. *Science* **242**: 399–405.
- White AG, Lespinasse Y. 1986.** The inheritance of fruit colour in apple (*Malus pumila* Mill.). *Agronomie* **6**: 105–108.
- Wilson DN, Cate JHD. 2012.** The structure and function of the eukaryotic ribosome. *Cold Spring Harbor Perspectives in Biology* **4**: 5.

- Wright SI, Agrawal N, Bureau TE. 2003.** Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Research* **13**: 1897–1903.
- Wu H, Qin X, Wu H, Li F, Wu J, Zheng L, Wang J, Chen J, Zhao Y, Lin S, et al. 2020.** Biochar mediates microbial communities and their metabolic characteristics under continuous monoculture. *Chemosphere* **246**: 125835.
- Wyatt GR. 1951.** The purine and pyrimidine composition of deoxypentose nucleic acids. *The Biochemical journal* **48**: 584–590.
- Xie Z, Johansen LK, Gustafson AM, Kasschau KD, Lellis AD, Zilberman D, Jacobsen SE, Carrington JC. 2004.** Genetic and functional diversification of small RNA pathways in plants. *PLoS Biology* **2**: 642–652.
- Xing L, Li Y, Qi S, Zhang C, Ma W, Zuo X, Liang J, Gao C, Jia P, Shah K, et al. 2019.** Comparative RNA-Sequencing and DNA Methylation Analyses of Apple (*Malus domestica* Borkh.) Buds with Diverse Flowering Capabilities Reveal Novel Insights into the Regulatory Mechanisms of Flower Bud Formation. *Plant & cell physiology* **60**: 1702–1721.
- Xiong T, Meister GE, Workman RE, Kato NC, Spellberg MJ, Turker F, Timp W, Ostermeier M, Novina CD. 2017.** Targeted DNA methylation in human cells using engineered dCas9-methyltransferases. *Scientific Reports* **7**: 6732.
- Xu J, House M, Booker H, Tanino K. 2018a.** Exploiting epigenetic variation for plant breeding. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* **13**.
- Xu J, Zhou S, Gong X, Song Y, van Nocker S, Ma F, Guan Q. 2018b.** Single-base methylome analysis reveals dynamic epigenomic differences associated with water deficit in apple. *Plant Biotechnology Journal* **16**: 672–687.
- Yamamoto N, Kobayashi H, Togashi T, Mori Y, Kikuchi K, Kuriyama K, Tokuji Y. 2005.** Formation of embryonic cell clumps from carrot epidermal cells is suppressed by 5-azacytidine, a DNA methylation inhibitor. *Journal of Plant Physiology* **162**: 47–54.
- Yang G, Dong J, Chandrasekharan M, Hall T. 2001.** Kiddo, a new transposable element family closely associated with rice genes. *Molecular Genetics and Genomics* **266**: 417–424.
- Yang J, Gao M, Huang L, Wang Y, Van Nocker S, Wan R, Guo C, Wang X, Gao H. 2017.** Identification and expression analysis of the apple (*Malus × domestica*) basic helix-loop-helix transcription factor family. *Scientific Reports* **7**: 1–14.
- Yang X, Kundariya H, Xu YZ, Sandhu A, Yu J, Hutton SF, Zhang M, Mackenzie SA. 2015.** Muts HOMOLOG1-derived epigenetic breeding potential in tomato. *Plant Physiology* **168**: 222–232.
- Yarychkivska O, Shahabuddin Z, Comfort N, Boulard M, Bestor TH. 2018.** BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate de novo and maintenance methylation in vivo. *Journal of Biological Chemistry* **293**: 19466–19475.
- Yu A, Lepère G, Jay F, Wang J, Bapaume L, Wang Y, Abraham AL, Penterman J, Fischer RL, Voynet O, et al. 2013.** Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 2389–2394.

Zhang Q, Arbuckle J, Wessler SR. 2000. Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family Heartbreaker into genic regions of maize. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 1160–1165.

Zhang Y, Harris CJ, Liu Q, Liu W, Ausin I, Long Y, Xiao L, Feng L, Chen X, Xie Y, et al. 2018. Large-scale comparative epigenomics reveals hierarchical regulation of non-CG methylation in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **115**: E1069–E1074.

Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H, et al. 2019. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nature Communications* **10**.

Zhang Y, Lespinasse Y, Salesses G. 1988. Mise en évidence de quelques anomalies méiotiques conduisant à la formation de gamètes mâles non réduits chez le pommier cultivé (*Malus ×domestica* Borkh.). *Cytologia* **53**: 749–755.

Zhang XZ, Zhao YB, Li CM, Chen DM, Wang GP, Chang RF, Shu HR. 2007. Potential polyphenol markers of phase change in apple (*Malus domestica*). *Journal of Plant Physiology* **164**: 574–580.

Zhao Y. 2010. Auxin Biosynthesis and Its Role in Plant Development. *Annual Review of Plant Biology* **61**: 49–64.

Zhao D, Song GQ. 2014. Rootstock-to-scion transfer of transgene-derived small interfering RNAs and their effect on virus resistance in nontransgenic sweet cherry. *Plant Biotechnology Journal* **12**: 1319–1328.

Zhao Y, Zhang C, Liu W, Gao W, Liu C, Song G, Li WX, Mao L, Chen B, Xu Y, et al. 2016. An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. *Scientific Reports* **6**: 1–11.

Zhong X, Du J, Hale CJ, Gallego-Bartolome J, Feng S, Vashisht AA, Chory J, Wohlschlegel JA, Patel DJ, Jacobsen SE. 2014. Molecular Mechanism of Action of Plant DRM De Novo DNA Methyltransferases.

Zimmerman RH. 1973. Juvenility and flowering of fruit trees. In: Acta Horticulturae 34: Symposium on Growth regulators in fruit production. Beltsville, Maryland, USA, 139–142.

Annexes

Annexes à l'article (figures supplémentaires)

The following supplemental material is available for this article:

Accession numbers:

GSE138492: global depository accession number comprising methylome and transcriptome data

GSE138377: bisulfite sequence data and methylation calling files + Supplemental table S3

GSE138491: microarray data

A

Gene_ID	Ratio		TAIR name	Short description	Forward primer	Reverse primer
	OGvsSD	OGvsYG				
MD01G1090900	4.51		AT1G01060	Homeodomain-like superfamily protein	CAGAGTTTTTCACCGCCTTC	CTCGTGCTGGTGACACACTT
MD02G1302200	3.17	2.11	AT3G63490	Ribosomal protein L1p/L10e family	CAGAATTC AAGCAGGGGAAA	CACCTGGTGGCTTATTTGCT
MD03G1128600	3.09	1.91	AT3G44890	ribosomal protein L9	GCTGCTCAAGGAAATGAAGG	TCAACGAGATCTTGGGGTGT
MD12G1237700	3.01	2.76	ATCG01120	chloroplast ribosomal protein S15	GAAGAAAAAGACGAAAAACAGAGG	TGACTCCCGAATATCCAACC
MD16G1267700		3.01	AT5G14460	Pseudo-uridine synthase family protein	CCTGCAAAGAAAAGCAAAGG	CTCTTATCAGCCTCGCCATC
MD17G1271400	3.2		AT5G17300	Homeodomain-like superfamily protein	GAACAGCCAGTGATGACGAA	TACTTCCGGTGCTTGCTCT
MD02G1099300	-1.75	-3.2	AT5G20790	Transmembrane protein	CATTGAGTTTCCGCTCACAA	CCAGGTCAACGTCACCTCA
MD09G1006500	-3.81	-3.72	AT3G02040	Senescence-related gene 3	CCGCTCTCCAAGTGGTGTTT	GCCTTTAACCTCCGACACAA

B

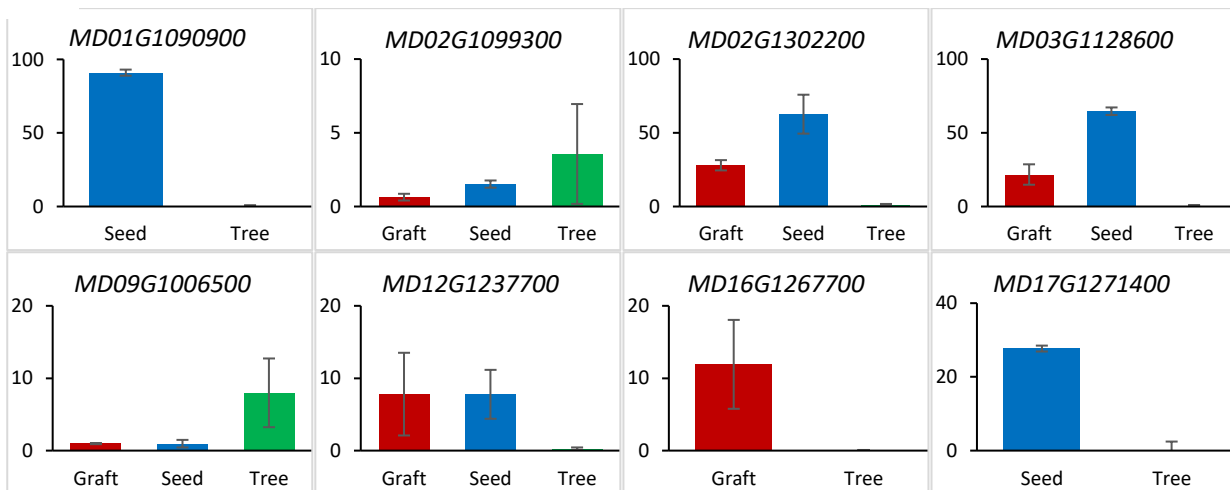


Fig. S1 Microarray validation. (A) Q-PCR primers for micro array data validation. Indicated ratios came from micro array data in both comparisons. (B) Q-PCR validation of micro array data.

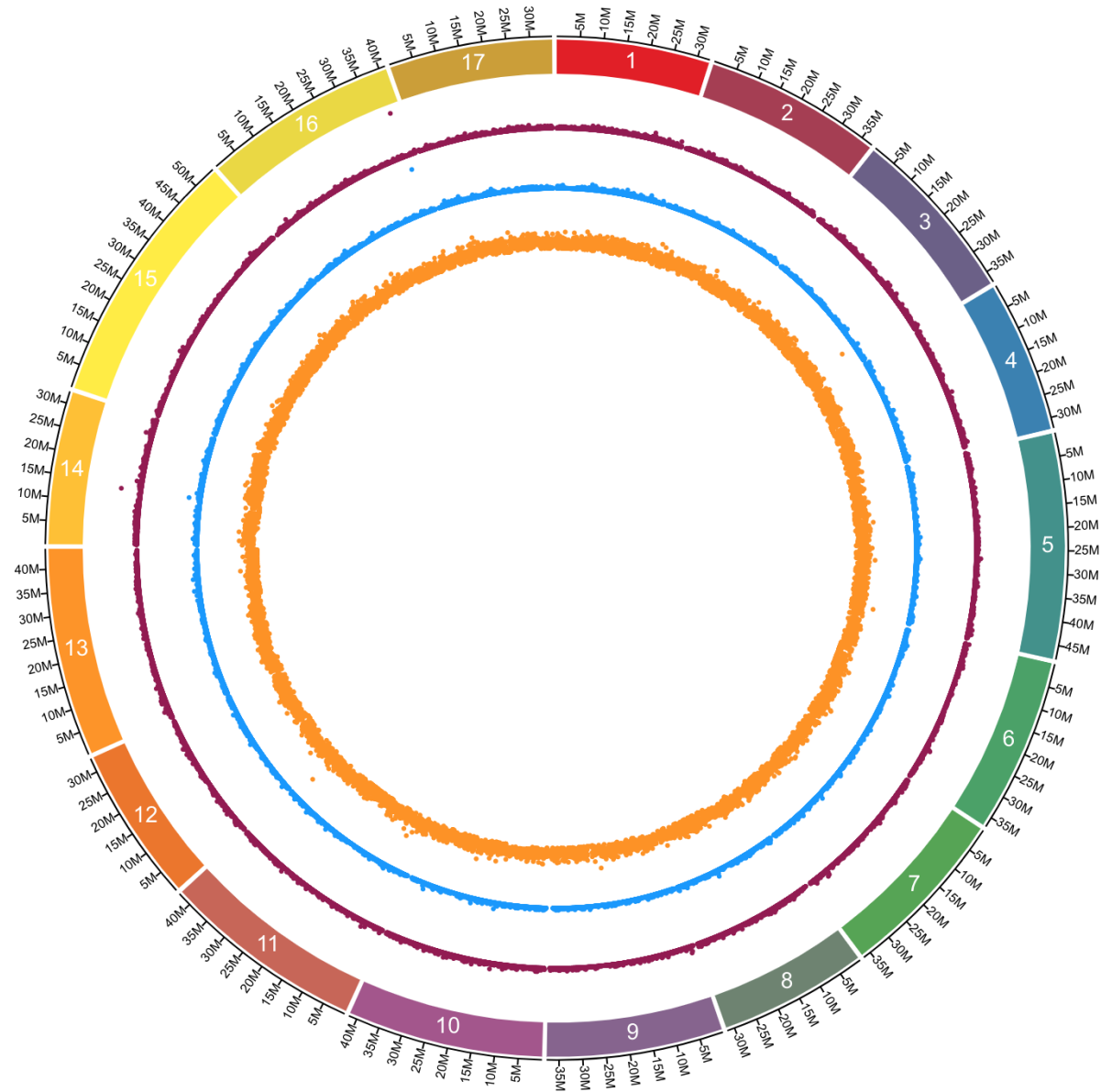


Fig. S2 Methylation overview in GDDH13. Density plot of DMRs on all GDDH13 genome in OGvYG. In red, DMRs in CG context, in blue CHG and in orange CHH. Each point represent number of DMRs in 50kb windows of genome.

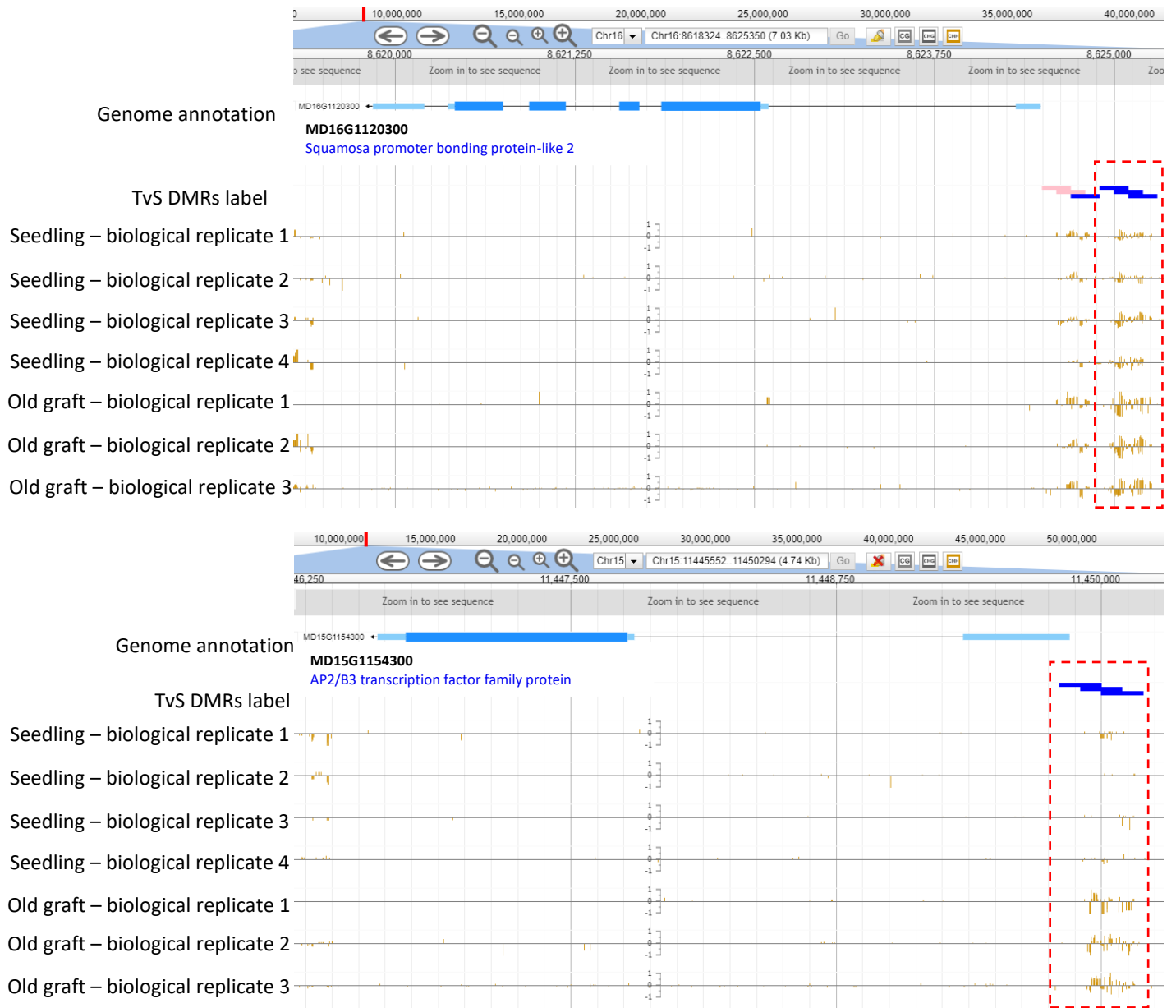


Fig. S3 Jbrowse screenshot of two DTG-DMRs present in scatterplot (fig 7A) with DMRs in the promotor of genes highlight by a red dashed boxe.

Table S1 Resume of defined samples and details of sampling

Sample	Way of multiplication	Years of obtention	Years sampling	Numbers sampled plants	Numbers leaves sampled per plant	Pooled	Number of biological replicates
Old graft	Grafting (Asexual)	2005	2016	1	4	Yes	3
Seedling	Seedling (Sexual)	2016	2016	4	1	No	4
Young graft	Grafting (Asexual)	2016	2016	20	1	Yes	2

Table S2 Fixed threshold to filter DMRs calculated between each comparison. Threshold were empirically fixed by observation of methylation calling file in the Jbrowse software.

DMRs	Standard deviation average threshold			p-value
	Old graft	Seedling	Young graft	
CG -CHG	< 0,07	< 0,11	< 0,07	< 1%
CHH	< 0,05	< 0,05	< 0,05	

Table S3: DMRs list of OGvsSD and OGvsYG comparisons, including Gene- and TE-DMRs. File deposited on GEO with the following accession number GSE138377

Table S4: DETs list of OGvsSD and OGvsYG comparison. Excel file available in additional information (named: “Supp_Tab_S4”)

Table S5: DTG- and DTTE-DMRs list of OGvsSD and OGvsYG comparisons. Excel file available in additional information (named: “Supp_Tab_S5”)

Table S6 Functional classes found in enrichment analysis on Mapman software. In blue are indicated the number of DTGs up-regulated in SD or YG compare to OG, down-regulated ones are in red. “p-value” corresponds to the p-value obtained in the enrichment analysis and corrected by the BH method.

Category name	Bincode	OGvsSD			OGvsYG			Commons			
		Down	Up	p-value	Down	Up	p-value	Down	Up	other	p-value
Coenzyme metabolism	7	0	19	2.56E-05				1	3		1.59E-02
Secondary metabolism / terpenoids	9.1	2	11	5.43E-03							
SECONDARY METABOLISM / PHENOLICS / P-COUMAROYL-COA SYNTHESIS	9.2.1				0	2	7.50E-03				
Chromatin organization	12	19	3	2.48E-05	1	14	6.03E-03				
Cell cycle	13				5	39	9.05E-04				
Rna biosynthesis / transcriptional activation / SBP transcription factor	15.7.18	8	0	3.37E-03							
Rna biosynthesis / organelle machineries	15.9							1	5		1.50E-02
RNA processing	16				5	30	3.53E-03				
Protein biosynthesis	17	9	40	3.17E-03	5	29	2.90E-04	5	53		4.07E-12
Protein modification / phosphorylation / TKL kinase superfamily	18.8.1							1	9	1	6.55E-03
Protein degradation / peptide tagging	19.4	18	15	7.94E-03	16	7	5.10E-04				
Protein degradation / peptidase families / serine-type peptidase activities	19.5.2							1	9	1	8.33E-04
Cytoskeleton	20				1	20	1.04E-06				
Cell wall	21				13	0	1.58E-05				
Protein translocation / chloroplast / thylakoid membrane SRP insertion system	23.1.7							1	7	1	1.50E-02
Solute transport	24				16	1	3.12E-06				
Enzyme classification	50	30	84	4.24E-03	31	25	7.32E-03				
Not assigned	35				235	617	7.50E-03	122	252	14	1.17E-02

Table S7 Count of DTG-DMRs in OGvsSD and OGvsYG comparisons. Here we included the unclassified gene class “35” not present in figure 6. Hypo- or hypermethylated in SD or YG compare to OG

Context	localization	OG vs SD			OG vs YG		
		Hypomethylated	Hypermethylated	Σ	Hypomethylated	Hypermethylated	Σ
CHH	head	1394	55	1449	406	261	667
	body	295	21	316	137	70	207
	tail	684	31	715	228	137	365
CHG	head	0	2	2	0	3	3
	body	1	5	6	0	2	2
	tail	0	4	4	0	3	3
CG	head	1	5	6	0	7	7
	body	3	4	7	0	4	4
	tail	1	4	5	0	10	10
	Σ	2379	131	2510	771	497	1268

Table S8 Count of DTTE-DMRs in CHH context in OGvsSD and OGvsYG comparisons. Hypo- or hypermethylated in SD or YG compare to OG

Context	localization	OG vs SD			OG vs YG		
		Hypomethylated	Hypermethylated	Σ	Hypomethylated	Hypermethylated	Σ
CHH	head	0	0	0	0	0	0
	body	115	22	137	153	113	266
	tail	10	1	11	5	6	11
CHG	head	0	0	0	0	0	0
	body	0	3	3	1	28	29
	tail	0	0	0	1	0	1
CG	head	0	0	0	0	0	0
	body	1	6	7	1	22	23
	Tail	0	0	0	0	1	1
	Σ	126	32	158	161	170	331

Annexes Chapitre I


Annexe 1 : protocole de fixation et d'inclusion d'échantillon biologique (Source : Imac – IRHS).

IMAC	SFR 4207 - Plateau Technique Mutualisé IMAC	Ref :
	Fixation Glutaraldehyde 4%	Date de création : 19/04/2013 Page 1 sur 2

- Objectif :** La fixation est l'étape essentielle dans la préparation des tissus pour l'observation microscopique pour préserver la structure des tissus aussi fidèlement que possible par rapport à l'état vivant. Les tissus doivent être fixés immédiatement après leur prélèvement.
- Lieu :** Plateau Technique Mutualisé IMAC
- Personnel habilité :** Toute personne ayant pris connaissance des consignes de manipulation auprès des responsables du plateau technique mutualisé IMAC : Marjorie Juchaux, Fabienne Simonneau
- Matériel :** Hotte aspirante, pompe à vide, dessiccateur, tube à hémolyse



Utilisation de la pompe à vide : infos pratiques. La pompe à vide est en fonctionnement de 7H30 à 19H30 les jours ouvrés. Le fonctionnement de la pompe est identifié par la lampe témoin rouge située au milieu du couloir labos. En dehors de ces créneaux voir Marjorie Juchaux ou Fabienne Simonneau.

- Réactifs spécifiques :**  Glutaraldéhyde 25 %. Préparation pour utilisation en microscopie électronique. Les flacons de glutaraldéhyde sont stockés dans la chambre -20°C. Le flacon en cours d'utilisation doit être stocké à 4°C. La glutaraldéhyde est très volatile : prendre soin de "filmer" le bouchon du flacon.

- Solutions :** Tampon phosphate pH 7.2 (0.2 mol/L)

- **Solutions mères :**

Tampon A:

1.2 g Na₂ H PO₄, 2H₂O / 100 ml H₂O

Ou 2.4 g Na₂ H PO₄, 12H₂O / 100 ml H₂O

Tampon B:

1.02 g Na H₂ PO₄, 2H₂O / 100 ml H₂O

Ou 0.90 g Na H₂ PO₄, H₂O / 100 ml H₂O

- **Tampon A + B** (se conserve une semaine environ à 4°C)
40 ml Tampon A + 10 ml Tampon B. Ajuster à pH 7.2

- **Solution de glutaraldéhyde 4% : la solution de glutaraldéhyde est très volatile, elle doit être préparée extemporanément.**

1 ml de glutaraldéhyde + 6 ml tampon phosphate pH 7.2

- Organisation :** Réserver une plage horaire de préparation-fixation échantillon sur le planning du Plateau Technique Mutualisé IMAC

Rédigé ou révisé le : 19/04/2013

par: M Juchaux

Visa:

Approuvé le : 18/04/2013

par: S. Simonneau

Visa:

Validation Qualité le : 19/04/2013

par: M. Juchaux

Visa:

IMAC	SFR 4207 - Plateau Technique Mutualisé IMAC	Ref :
	Fixation Glutaraldehyde 4%	Date de création : 19/04/2013 Page 2 sur 2

8. Méthodologie :

- Les échantillons sont plongés dans la solution de glutaraldéhyde 4%. *Le volume de fixateur doit être égal à 50 fois le volume de l'échantillon*
- La fixation "standard" se fait à 4°C. Les tubes sont maintenus à 4°C sur la glace pilée
- Pour certains tissus plus fragiles (ex : méristèmes) et afin d'éviter les chocs thermiques la fixation peut se faire à température ambiante. Immerger les échantillons immédiatement après leur prélèvement. et si nécessaire disséquer dans la solution du glutaraldéhyde, sous hotte
- Mettre les tubes dans la cloche à vide (**ne pas oublier de retirer les bouchons !!**). Ouvrir le vide avec précaution et commencer à dégazer les échantillons. (Certains types d'échantillons flottent, essayer de les maintenir immerger à l'aide de grilles)
- Rompre le vide toutes les 4 mn environ. Tapoter les tubes sur le bord du dessiccateur pour faciliter le dégazage
- Après environ 2 h de vide changer la solution de glutaraldéhyde 4% des tubes
- Laisser 12 h à 4°C
- Rincer 2 fois avec du tampon phosphate pH 7.2 et stocker au réfrigérateur
- Stockage : veiller à changer régulièrement le tampon. : 2 à 3 fois par semaine si oxydation apparaît

9. **Risque et sécurité** : le glutaraldéhyde est très volatile, dangereux par contact cutané, oculaire et muqueuses. Manipuler la glutaraldéhyde avec des gants et sous la hotte

10. **Tri des déchets** : Glutaraldéhyde : poubelle déchets toxiques

11 **Sources** : Plateau Technique Mutualisé IMAC.

IMAC	SFR 4207 - Plateau Technique Mutualisé IMAC	Ref :
	Inclusion en résine Technovit 7100	Date de création : 16/04/2013 Page 1 sur 2

1. **Objectif** : Inclure dans de la résine Technovit 7100 des échantillons préalablement fixés au glutaraldéhyde 4%
2. **Lieu** : Plateau Technique Mutualisé IMAC
3. **Personnel habilité** : Toute personne ayant pris connaissance des consignes de manipulation auprès des responsables du plateau technique mutualisé IMAC : Marjorie Juchaux, Fabienne Simonneau
4. **Matériel** : Hotte aspirante, platine chauffante, agitateur magnétique, moules à inclusion, étuve (37°C), pompe à vide, dessiccateur
5. **Réactifs spécifiques** : Pack Résine Technovit 7100 (résine + durcisseur N° 1 et 2)
6. **Solutions** : Alcools : 50°- 70°- 80°- 95°- 100°
7. **Organisation** : Réserver une plage horaire de préparation-fixation échantillon sur le planning du Plateau Technique Mutualisé IMAC. Le protocole se déroule sur 3 jours



Utilisation de la pompe à vide : infos pratiques. La pompe à vide est en fonctionnement de 7H30 à 19H30 les jours ouvrés. Le fonctionnement de la pompe est identifié par la lampe témoin rouge située au milieu du couloir labos. En dehors de ces créneaux voir Marjorie Juchaux ou Fabienne Simonneau.

8. Méthodologie :

JOUR 1 : DESHYDRATATION ET PRE-INFILTRATION DES ECHANTILLONS

- Rincer les échantillons 3 fois à l'eau distillée. (Pipeter doucement pour ne pas abîmer les échantillons)
- Déshydratation à température ambiante :

Alcool 50°	1 bain de 10 minutes
Alcool 70°	1 bain de 10 minutes
Alcool 80°	1 bain de 10 minutes
Alcool 95°	1 bain de 10 minutes
Alcool 100°	3 bains de 15 minutes

Nota : Si nécessité absolue : les échantillons peuvent être stockés 1 à 2 heures dans le 2^{ème} alcool 100°

- Pré-infiltration : préparer la solution de pré-infiltration : alcool 100° / résine Technovit 7100 (grande bouteille) (V/V) et remplacer le dernier bain d'alcool 100° par la solution de pré-infiltration
- Placer le portoir de tubes sur la glace pilée dans la cloche à vide
- Appliquer le vide pendant 2 h
- Stocker les tubes 12 h à 4°C

Rédigé ou révisé le : 16/04/2013

par: M. Juchaux

Visa:

Approuvé le : 16/04/2013

par: F. Simonneau

Visa:

Validation Qualité le : 16/04/2013

Par M. Juchaux

Visa:

IMAC	SFR 4207 - Plateau Technique Mutualisé IMAC	Ref :
	Inclusion dans la résine Technovit 7100	Date de création : 16/04/2013 Page 2 sur 2

JOUR 2 : INFILTRATION (sous hotte)

- Infiltration : préparer la solution d'infiltration : dissoudre 1 sachet de durcisseur I (poudre) dans 100 mL de résine Technovit 7100. Agiter 5 à 10 minutes. Cette solution peut se conserver 1 mois à 4°C
- Remplacer la solution de pré-infiltration par la solution d'infiltration
- Placer le portoir de tubes sur la glace pilée dans la cloche à vide
- Appliquer le vide pendant 20 mn minimum. Le temps varie en fonction de la dureté de l'échantillon
- Placer les tubes 12 h à 4°C

JOUR 3 : INCLUSION (sous hotte, *attention: la température de l'air aspiré (froid l'hiver) influe sur la vitesse de polymérisation de la résine*)

- Inclusion : préparer la solution d'inclusion : 1ml de durcisseur II (liquide) (*petite bouteille*) et 15 ml de solution d'infiltration. *Préparer le volume de solution d'inclusion adéquat (moules en téflon classique environ 700 µL /alvéole*
- Mettre sous agitation assez vive pour éviter la polymérisation rapide de la résine
- Recouvrir la platine chauffante de papier aluminium; Allumer la platine, elle doit être tiède. 35°C -38°C environ. *L'utilisation d'une platine chauffante permet de régulariser la polymérisation de la résine*
- Mettre 250-300µl de solution dans chaque alvéole
- Vérifier à l'aide d'une aiguille le début de la polymérisation quand la résine commence à filer (de 10 à 30 mn en fonction de la T°)
- Quand la résine commence à filer : retirer l'excédent de solution d'infiltration puis enrober les échantillons avec de la solution d'inclusion
- Disposer et orienter les échantillons dans les alvéoles sans oublier de faire un relevé précis de l'emplacement des échantillons dans les moules

Ces dernières étapes doivent se faire rapidement. Le processus de polymérisation peut être très rapide

- Remplir de solution d'inclusion les alvéoles jusqu'à former un dôme bombé. *La résine se rétracte en séchant*
- Laisser polymériser 2 h environ
- Envelopper les moules dans du film étirable puis les placer dans l'étuve à 37°C afin de terminer la polymérisation pendant 12 h
- Démouler les échantillons. Les stocker à l'étuve à 37°C. Les coupes peuvent être réalisées après une semaine à l'étuve à 37 °C

9. Risque et sécurité : Les manipulations s'effectuent sous hotte avec des gants. La résine Technovit 7100 est irritante lorsqu'elle n'est pas polymérisée

10. Tri des déchets : Les solutions contenant de la résine sont déposées dans le flacon "RESINE" situé sous la hotte. Les autres solutions sont à déposer dans la poubelle déchets toxiques

11. Sources : Plateau Technique Mutualisé IMAC et notice d'utilisation fournisseur

Annexe 2: Nombre de DMRs calculées avec la première méthode de calcul, sur la V1.0 du génome de GDDH13

Bp in	context	TvS						TvG						GvS																			
		Hyperméthylé dans l'adulte			Hypométhylé dans l'adulte			Hyperméthylé dans l'adulte			Hypométhylé dans l'adulte			Hyperméthylé dans le greffon			Hypométhylé dans le greffon																
		Head	Body	Tail	Head	Body	Tail	Head	Body	Tail	Head	Body	Tail	Head	Body	Tail	Head	Body	Tail														
100	CG	34	119	16	14	55	4	NA	22	51	17	2	12	4	7	49	3	5	34	2													
	5mC	12	10	8	52	18	3														13	13	11	68	20	18	25	41	28	18	18	15	5
	CHG	33	131	15	12	52	9														13	55	20	14	10	25	8	9	43	3	6	50	2
	5mC	27	36	31	42	32	24														22	24	26	25	38	33	43	44	38	29	22	34	7
	CHH	2	35	4	2	14	2														1	11	1	1	4	19	6	0	9	0	6	48	2
	5mC	55	37	45	10	18	15														19	43	49	16	23	28	25	/	23	/	36	26	28
200	CG	58	127	30	17	57	8	NA	52	51	17	2	12	4	17	49	10	15	34	7													
	5mC	15	10	7	41	20	9														13	13	11	68	20	18	25	28	15	33	15	14	
	CHG	67	135	22	19	55	14														13	55	20	14	10	25	8	13	43	9	17	50	7
	5mC	28	36	29	41	32	24														22	24	26	25	38	33	43	38	30	35	34	34	21
	CHH	4	35	6	1	15	2														1	11	1	1	4	19	6	3	9	2	14	48	10
	5mC	59	37	34	10	17	16														19	43	49	23	28	28	25	23	12	39	26	26	23
500	CG	141	127	84	37	57	22	NA	52	51	27	10	12	6	40	49	17	33	34	14													
	5mC	13	10	10	31	20	13														11	13	11	23	20	13	19	28	19	31	15	15	
	CHG	147	135	86	43	55	32														48	55	30	32	27	25	16	34	43	13	53	50	24
	5mC	27	36	29	32	32	25														23	24	26	29	38	33	43	33	38	29	31	34	32
	CHH	8	35	9	4	15	2														1	9	1	1	13	18	6	4	9	2	35	48	21
	5mC	52	37	33	28	17	15														19	42	49	23	29	22	29	27	23	12	31	26	25
1000	CG	276	127	161	66	57	49	NA	108	51	76	19	12	14	71	49	36	57	34	35													
	5mC	11	10	9	26	20	12														11	13	9	18	20	15	20	28	17	29	15	16	
	CHG	274	135	183	83	55	63														78	55	73	45	25	41	50	43	20	91	50	49	49
	5mC	26	36	29	31	32	26														23	24	24	33	38	37	31	38	30	27	34	29	29
	CHH	19	35	15	7	15	7														4	11	2	2	24	19	13	6	9	4	54	48	40
	5mC	45	34	35	26	17	23														37	43	58	24	23	25	29	23	21	29	26	26	26
2000	CG	436	127	308	114	57	84	NA	192	51	130	37	12	30	124	49	64	87	34	66													
	5mC	10	10	9	22	20	12														10	13	9	22	20	15	19	28	19	23	15	14	
	CHG	448	135	348	136	55	112														134	55	114	79	25	64	85	43	49	142	50	88	88
	5mC	27	36	29	29	32	26														22	24	24	35	38	39	28	38	29	25	34	27	27
	CHH	36	35	28	15	15	11														7	11	5	5	39	19	22	8	9	8	98	48	76
	5mC	39	37	36	21	17	20														48	43	45	25	23	28	21	23	26	27	26	26	26

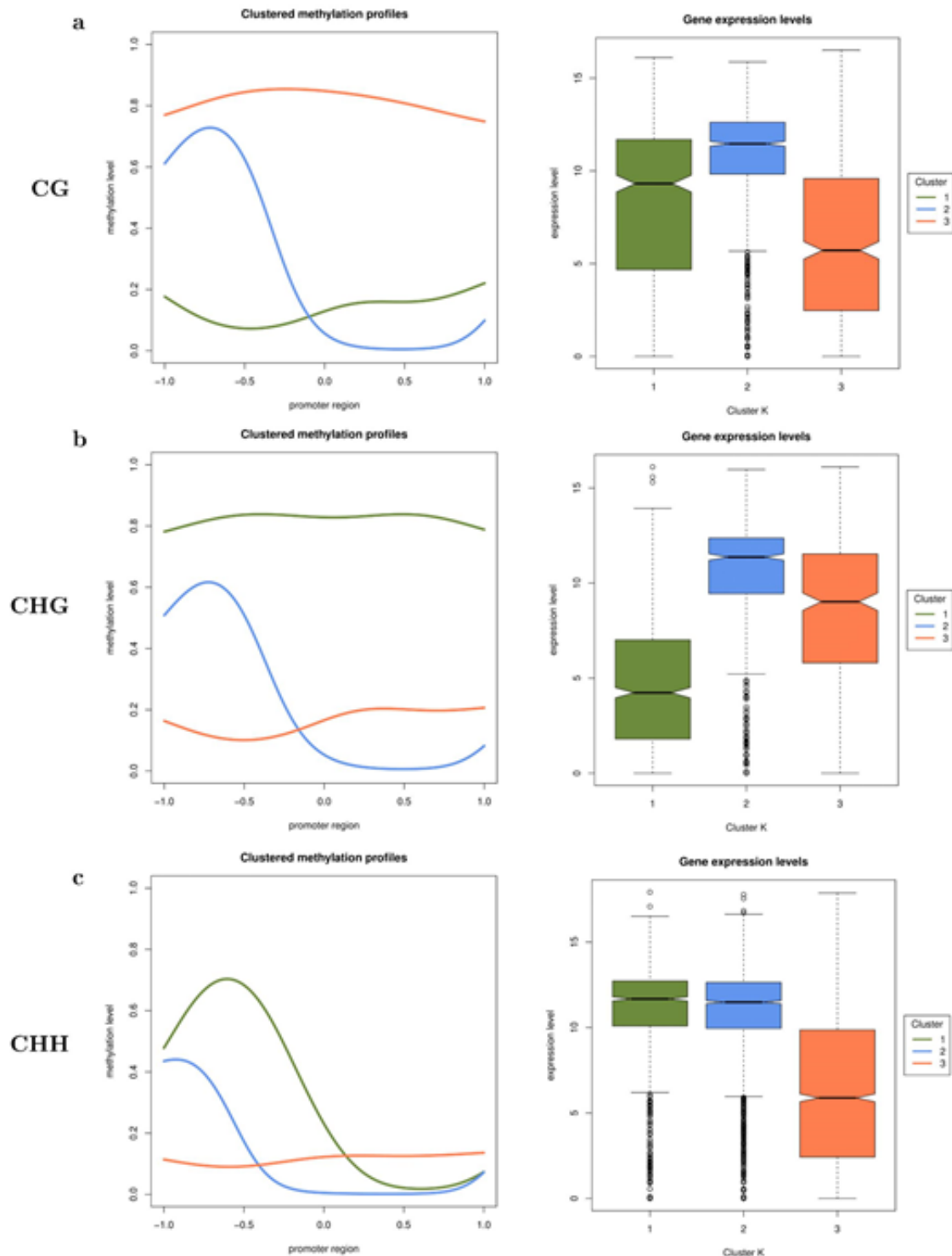


Figure 3.4 Results of the methylation patterns analysis, for the [-100 :100] regions with three clusters of patterns. **(a (left))** Line plot of the methylation patterns for the CG methylation context. The y-axis represents the methylation ratio and the x-axis represent the position relatively to the TSS. On the x-axis, -1.0 represents the nucleotide which is one hundred bases before the TSS, 0.0 represents the TSS position and 1.0 represents the nucleotide which is one hundred bases after the TSS. Each line corresponds to a different cluster. **(a (right))** Boxplot representing the expression values of the genes of each clusters represented in (a (left)). The boxplot colors are corresponding to the line colors in (a (left)). **(b)** Line plot and boxplot expression values for the CHG methylation context. **(c)** Line plot and boxplot expression values for the CHH methylation context.

Annexes Chapitre II

Annexe 4 : Protocole l'électroporation d'Agrobacterium tumefaciens compétentes utilisé. Protocole provenant de l'IRHS (ANGERS).

	Institut de Recherche en Horticulture et Semences Centre INRA Angers-Nantes, 42 rue Georges Morel, 49071 Beaucouzé	Ref : IRHS-MO-RR-0308
	Electroporation d'agrobacterium tumefaciens compétentes	Version : A Date de création : 10.07.2015 Page : 1 sur 2

Ref: IRHS-T-0001 / Version: A / Date de création: 13/03/2015

Objectif :	transférer un plasmide d'intérêt dans une souche d'agrobacterium
Lieux :	Laboratoires de bactériologie« OGM » et de biologie moléculaire
Personnel habilité :	Toute personne ayant pris connaissance des consignes de manipulation auprès de la personne compétente au sein de son équipe

Références :

Matériel :

Matériel classique de biologie moléculaire et de bactériologie.

Réactifs spécifiques et Solutions :

Miniprep du plasmide d'intérêt
Agrobacterium électro-compétentes
RMQ : pour le pommier et le Poirier il s'agit de la souche EHA105 disposant du plasmide ternaire ppbr-MCS5 qui porte un gène de résistance à la gentamycine
Boîte de Pétri milieu Luria-Bertani (LB)-agar + Antibiotiques (celui correspondant au gène de résistance porté par le plasmide d'intérêt + gentamycine 50 µg/mL final).

Risques et sécurité :

Manipuler les bactéries sous PSM.

Tri des déchets :

Les « plastiques » (cônes, spatules et râpeaux...) entrés en contact avec les bactéries compétentes doivent être jetés dans les poubelles déchets biologiques, la verrerie et les résidus de culture doivent être autoclavés.

Organisation :

- Prévoir les poubelles correspondantes aux déchets.
- Utiliser des gants pour les manipulations hors PSM.
- Réserver le PSM.

Méthodologie :

- Sortir les agrobacterium compétentes du congélateur -80°C et les placer tout de suite sur glace. Un tube contient 100 µl de bactéries compétentes.
- Les laisser décongeler tranquillement sur glace.
- Régler les paramètres de l'électroporateur (Multiporator, Eppendorf) pour agrobacterium : tension 1500 V, τ

Rédigé ou révisé le : 10.07.2015 Par : Emilie VERGNE	Vérifié le : 10.07.15 Par : Aurélie Charrier	Approuvé le : 10.07.15 Par : Emilie Vergne
Visa : 	Visa : 	Visa : 

	Institut de Recherche en Horticulture et Semences Centre INRA Angers-Nantes, 42 rue Georges Morel, 49071 Beaucouzé	Ref : IRHS-MO-(type d'activité)-0001
	<h2 style="color: green;">Clonage réaction LR</h2>	Version : A Date de création : 13.03.2015 Page : 2 sur 2

- (constante de temps) 5 ms.
- Ajout de 1 à 3 µL de plasmide d'intérêt (150 ng, la miniprep doit être la plus propre possible (dosage au nanodrop) dans un tube de bactéries.
 - Transférer dans une cuve d'électroporation 1 mm.
 - Placer la cuve dans l'électroporateur, LAISSER LE CAPUCHON.
 - Appuyer sur « start ».
 - Noter les paramètres qui s'affichent pour le choc appliqué : tension et constante de temps
 - Sous PSM ajouter 1 ml de LB à température ambiante, SANS ANTIBIOTIQUES.
 - Incubation 1 à 2h, à 28°C sous agitation.
 - Etaler 50 et 500 µl sur deux boîtes LBA + antibiotiques
 - Incuber 48h maximum à 28°C
 - Sélectionner et isoler 4 clones transformés sur de nouvelles boîtes LBA + antibiotiques (diviser la boîte en 4 quarts).
 - Effectuer l'analyse des clones par PCR ou digestion enzymatique.
 - Sélectionner un clone transformé pour le plasmide d'intérêt et en faire un stock glycérol
 - Placer le stock glycérol dans la banque à -80°C et remplir le fichier de suivi des souches

Rédigé ou révisé le : 10.07.2015 Par : Emilie VERGNE	Vérifié le : 10.07.2015 Par : Aurélie Charrier	Approuvé le : 10.07.2015 Par : Emilie VERGNE
Visa : 	Visa : 	Visa : 

Annexe 5: composition de différents milieux de culture utilisés pour la culture *in vitro* du pommier. Composition adapté de Murashige and Skoog (1962). * on peut employer la poudre MS pour l'entretien des collections. ** on peut employer du sucre ordinaire pour l'entretien des collections

Milieu de micropropagation POM	"classique"	"classique"	variante "L"	variante "D"	variante "A"
Quantités pour 1 l de milieu	concentration finale en mg/l	à partir de poudre MS*	milieu L à partir de solutions mères	milieu D à partir de solutions mères	milieu A à partir de solutions mères "milieu plantes albinos"
Macro éléments MS (Murashige et al., 1962)					
KNO ₃	1900	4,4 g/l	50 ml	50 ml	50 ml
NH ₄ NO ₃	1650				
MgSO ₄ ,7H ₂ O	370				
KH ₂ PO ₄	170				
CaCl ₂ .2H ₂ O	440				
Micro éléments MS (Murashige et al., 1962)					
MnSO ₄ ,H ₂ O	22.3	4,4 g/l	1 ml	1 ml	1 ml
ZnSO ₄ ,7H ₂ O	8.6				
H ₃ BO ₃	6.2				
KI	0.83				
CoCl ₂ ,6H ₂ O	0.025				
CuSO ₄ ,5H ₂ O	0.025				
Na ₂ MoO ₄ ,2H ₂ O	0.25				
EDFS (=Fer-EDTA)	42.11		5 ml	5 ml	5 ml
Vitamines et acides aminés MS (Murashige et al., 1962)					
Myo-inositol	100	4,4 g/l	10 ml	10 ml	10 ml
Pyridoxine	0.5				
Acide nicotinique	0.5				
Thiamine	0.1				
Glycine	2				
BAP	0.5	0,5 ml	0,5 ml	0,5 ml	0,5 ml
AIB	0.1	0,1 ml	0,1 ml		0,1 ml
	g/l	g/l	g/l	g/l	g/l
Saccharose**	30	30	30		
Sucrose					15
Glucose				15	
Fructose				15	
Sorbitol					15
European microbiological agar (labo CONDA)	7.5	7.5	7.5	7.5	7.5
Ajustement du pH	5.75				

Annexe 6: liste des amorces de la thèse

Numéro de l'amorce	Nom de l'amorce	Séquence de l'amorce	Génome source Ou source	Gène ou transposon ciblé	
1	CRISPR_MET1_MD02-F	GGGGGGGAGTACTATTTAAATAAT	GDDH13	MD02G1115100	
2	CRISPR_MET1_MD02-R	CGCTCCATCAATCTCCATGG			
3	CRISPR_MET1_MD15-F	TGAGGAGGAAGCTAATGGTGA	GDDH13	MD15G1231100	
4	CRISPR_MET1_MD15-R	CTTGTTTCATACATATACTCCACAT			
5	pHODOR_F	CCCAAAGGTTGCAAGGTAGTG	GDDH13	HODOR	
6	pHODOR_R	AACCGAGAGTGATCTTCGGTG			
7	pLTR-HODOR_F	CTTCGAACGTC AAGCCATTG			
8	pLTR-HODOR_R	CCCAAAGAGAGGGCCTTCATT			
9	EF1 α -F	CTCTGGGTGCAGGCAAAT	M. domestica Borkh, Gala Respom	AJ223969	
10	EF1 α -R	TCAAGGTTGGTGGACCTCT			
13	Agro_F (UF)	GTAAGAAGCGAACGCAGGGAACCT	Agrobacterium tumefaciens Equipe Respom	sous-unité 23S de l'ARN ribosomique d'Agrobacterium tumefaciens	
14	Agro_R (B1R)	GACAATGACTGTTCTACGCGTAA			
15	NPTII_F (owb24)	GAGGCTATTCGGCTATGACTG	Equipe Respom	gène de résistance à la kanamycine	
16	NPTII_R (owb25)	ATCGGGAGCGGCATACCGTA			
17	pDECas9Kr-F	AGCTCCTAGGCTGTTATC	Equipe Respom	Vérification de l'insertion de la cassette des gRNA dans pDECas-kr par gateway	
18	pDECas9Kr-R	CTAGGCTGGATCGGAATTATCG			
19	pDONR207_F	TCGCGTTAACGCTAGCATGGATCTC	Equipe Respom	Vérification de l'insertion de la cassette des gRNA dans pDonR207 par gateway	
20	pDONR207_R	GTAACATCAGAGATTTGAGACAC			
21	Cas9_F	ACCACTCAGAAAGAAGCTCG	pDECas9-kr	Vérification de l'insertion de la Cas9 <i>in planta</i>	
22	Cas9_R	CCTTGGTGATCTCGGTGTTCA		Modification des nucléotides CA en GC en position 2518 de la Cas9	
23	Cas9-840_F	GATTACGATGTTGATCATATCGTGCCACAGTC		Modification du nucléotide A en C en position 28 de la Cas9	
24	Cas9-840_R	GACTGTGGCAGATATGATCAACATCGTAATC		Amplification Fragment 1	
25	Cas9-10_F	TTCGGCGGCCATGGATAAGAAGTACTCTATCGGACTCGATATCGG		Amplification Fragment 2	
26	Cas9-10_R	CCGATAGCGAGTCCGATAGAGTACTTCTATCCATGGCGCCGAA		DeadCas9	
27	Cas9_840_Bsal_R	CGATGTTGATCATATCGTGCCACAGTCATT		Construction synthétisé HSP-DRM2	
28	Cas9_840_Bsal_F	TGTGGCACGATATGATCAACATCGTAATCAG			
29	Cas9_End_R	CTCGGCGCGCTCAAACCTTC		pDECas9-kr	DeadCas9
30	DeadCas9_Bsal_F	CGGTCTCCATGGATAAGAAGTACTCTATCGG			
31	HSP-DRM2_F	CGGTACCCAATTGTTAAGAG	Vérifier séquence de la deadCas9 et mutation des nucléases		
32	HSP-DRM2_R	GAATTCGGGTCTCGCCATAC			
33	DeadCas9_EcoRI_F	TTTGAATTCATGGATAAGAAGTACTCTATCGG			
34	Cas9_1_F	CATGGATAAGAAGTACTCTATCGG			
35	Cas9_2_R	TTGAGCTTCACGAGGAGTTCC			
36	Cas9_3_F	ATCGATGGTGGTGCATCTCAA			
37	Cas9_4_R	CCCTTCTGAGTGGTCTGGTTC			
38	Cas9_5_F	TGAGTTGGTGAAGGTGATGGG			
39	Cas9_6_R	TAACGATGTTACCTGTGGCA			
40	Cas9_1_F	CGAGACAAACGGTGAACAGG			
41	dCas9_Swal-Apal_F	AAAATTTAAATTTGGGCCATGGATAAGAAGTACTCTAT		Réalisation du protocole figure 3.2.2	
42	Swal-HSP-DRM2_F	AAAATTTAAATTTAAGAGATAGAATTTAAA			
43	HSP-DRM2-Apal_R	TTTGGGCCGAAAGGCTATTGGAAGAGA			
44	Ligation_DRM2-deadCas9_F	GGTAAGCCGAGTTGGTCTTGA			
45	Ligation_DRM2-deadCas9_R	TAGCCTCAGCAGTTCCACCAG			

Annexe 7 : liste des méthyltransférases identifiées sur la V1.0 du génome de référence

MD_gene	TAIR-name	TAIR_evalue	TAIR_annot	TAIR_annot	e	evalue	Swissprot_annot	e
0	AT2G26810	1.00E-51	Putative_methyltransferase_family_protein	Putative_methyltransferase_family_protein	MET22_MOUSE	2.00E-09	like_protein_22_EC=2.1	MET22_MOUSE
0	AT2G31740	7.00E-168	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	N	4.00E-56	like_protein_13_EC=2.1	N
0	AT2G31740	6.00E-123	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET13_BOVIN	5.00E-41	like_protein_13_EC=2.1	MET13_BOVIN
0	AT4G28830	2.00E-113	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET15_MOUSE	8.00E-72	like_protein_5_EC=2.1	MET15_MOUSE
0	AT5G49160	2.00E-144	thyltransferase_1	thyltransferase_1	DNM1B_ORYSJ	1.00E-155	methyltransferase_1B_	DNM1B_ORYSJ
0	AT3G60910	4.00E-135	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET13_BOVIN	6.00E-25	like_protein_13_EC=2.1	MET13_BOVIN
0	AT2G31740	8.00E-52	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET13_DANRE	9.00E-28	like_protein_13_EC=2.1	MET13_DANRE
0	AT2G31740	0	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET13_BOVIN	4.00E-103	like_protein_13_EC=2.1	MET13_BOVIN
0	AT2G31740	0	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET13_BOVIN	4.00E-103	like_protein_13_EC=2.1	MET13_BOVIN
0	AT1G19340	1.00E-143	Methyltransferase_M1-A70_family_protein	Methyltransferase_M1-A70_family_protein	MET12_ARATH	2.00E-142	like_protein_2_EC=2.1	MET12_ARATH
0	AT1G54650	1.00E-132	Methyltransferase_family_protein	Methyltransferase_family_protein	MET1_DROME	4.00E-37	like_protein_EC=2.1.1.-	MET1_DROME
0	AT2G21070	0	F101;_methyltransferases	F101;_methyltransferases	MET16_DANRE	1.00E-67	like_protein_16_EC=2.1	MET16_DANRE
0	AT1G64600	0	methyltransferases;copper_ion_binding	methyltransferases;copper_ion_binding	MET17_MOUSE	1.00E-30	like_protein_17_mitoc	MET17_MOUSE
0	AT1G64600	0	methyltransferases;copper_ion_binding	methyltransferases;copper_ion_binding	MET17_MOUSE	1.00E-30	like_protein_17_mitoc	MET17_MOUSE
0	AT3G60910	2.00E-129	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET13_BOVIN	5.00E-29	like_protein_13_EC=2.1	MET13_BOVIN
0	AT5G04610	6.00E-155	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET13_XENLA	4.00E-27	like_protein_13_EC=2.1	MET13_XENLA
0	AT4G09980	0	A70 family protein	A70 family protein	MET11_ARATH	0	like_protein_1_EC=2.1	MET11_ARATH
0	AT2G26200	0	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET16_RAT	2.00E-51	like_protein_6_EC=2.1	MET16_RAT
0	AT3G01660	1.00E-128	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	N	0.00005	like_protein_7B_EC=2.1	N
0	AT1G58120	8.00E-61	is:_methyltransferases_(TAIR:AT5G01710.1);	is:_methyltransferases_(TAIR:AT5G01710.1);	RS19_METB6	0.57	_S19_{ECO:0000255}_H	RS19_METB6
0	AT4G34360	3.00E-134	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET13_DROME	1.00E-21	like_protein_13_EC=2.1	MET13_DROME
0	AT3G17365	3.00E-124	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET13_BOVIN	9.00E-28	like_protein_13_EC=2.1	MET13_BOVIN
0	AT4G13330	1.00E-35	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	N	0.0005	like_protein_13_EC=2.1	N
0	AT1G64600	0	methyltransferases;copper_ion_binding	methyltransferases;copper_ion_binding	MET17_MOUSE	8.00E-31	like_protein_17_mitoc	MET17_MOUSE
0	AT2G26200	2.00E-59	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET16_RAT	3.00E-17	like_protein_6_EC=2.1	MET16_RAT
0	AT2G26200	3.00E-09	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	MET16_HUMAN	0.002	like_protein_6_EC=2.1	MET16_HUMAN
0	AT4G14140	0	ansferase_2	ansferase_2	DNM1B_ORYSJ	0	methyltransferase_1B_	DNM1B_ORYSJ
0	AT1G63855	1.00E-83	Putative_methyltransferase_family_protein	Putative_methyltransferase_family_protein	MET23_XENLA	9.00E-42	like_protein_23_EC=2.1	MET23_XENLA
0	AT1G69523	3.00E-104	dependent_methyltransferases_superfamily	dependent_methyltransferases_superfamily	N	2.00E-36	like_protein_7A_EC=2.1	N
0	AT4G14140	4.9	ansferase_2	ansferase_2	Y227_ATV	0.94	n_ORF227	Y227_ATV
0	AT1G79915	2.00E-112	Putative_methyltransferase_family_protein	Putative_methyltransferase_family_protein	MET22_MOUSE	9.00E-35	like_protein_22_EC=2.1	MET22_MOUSE


Tableau 2

guideid	guideSeq	offtargetSeq	mismatchPos	mismatchCount	chrom	start	end	locusDesc
470forw	GAACGATAATGCGAGTAACGG	GAACGATATGCGAGTAACAG*	2	Chr07	2.1E+07	2.1E+07	intron:mRNA:MD07G1142700
470forw	GAACGATAATGCGAGTAACGG	GAAACATATACAGCAATAGG*	4	Chr01	1.5E+07	1.5E+07	intron:mRNA:MD01G1048700
470forw	GAACGATAATGCGAGTAACGG	GAAGGATGAAGCAGCAATAATGG*	4	Chr06	2.6E+07	2.6E+07	intergenic:mRNA:MD06G1120300-mRNA:MD06G1120400
470forw	GAACGATAATGCGAGTAACGG	TAAGGATAATGCGAGCAATGG*	4	Chr06	695382	695382	exon:mRNA:MD06G1006700
470forw	GAACGATAATGCGAGTAACGG	GAAAGGTAATGCGAGCAATAGG*	4	Chr03	8350861	8350883	exon:mRNA:MD03G1097100
470forw	GAACGATAATGCGAGTAACGG	GAAAGGTAATGCGAGCAATAGG*	4	Chr11	1.3E+07	1.3E+07	exon:mRNA:MD11G1143700
470forw	GAACGATAATGCGAGTAACGG	GAACGATGAGGAAAGAAAGGG*	4	Chr07	3420600	3420622	exon:mRNA:MD07G1040900
470forw	GAACGATAATGCGAGTAACGG	GAAACGATAATGCGAGTAACGG*	4	Chr11	3.2E+07	3.2E+07	intergenic:mRNA:MD11G1220300-mRNA:MD11G1220400
470forw	GAACGATAATGCGAGTAACGG	GACCGTACTGCAGCAATAGG*	4	Chr11	1.6E+07	1.6E+07	intron:mRNA:MD11G1160800
470forw	GAACGATAATGCGAGTAACGG	GAAAAAATGCGAGCAATAGG*	4	Chr13	2.6E+07	2.6E+07	intergenic:mRNA:MD13G1247900-mRNA:MD13G1248200
553rev	CCAGAACATGTTGGCTACAGG	CCAGAACATGTTGGCTACGG*	2	Chr14	2.6E+07	2.6E+07	intergenic:mRNA:MD14G1163300-mRNA:MD14G1163400
553rev	CCAGAACATGTTGGCTACAGG	CCAGAACATGTTGGCTACGG*	3	Chr06	1.9E+07	1.9E+07	intergenic:mRNA:MD06G1074900-mRNA:MD06G1075100
553rev	CCAGAACATGTTGGCTACAGG	CCAGAACATGTTGGCTACGG*	3	Chr16	1.5E+07	1.5E+07	intergenic:mRNA:MD16G1178900-mRNA:MD16G1179200
553rev	CCAGAACATGTTGGCTACAGG	CCAGAACATGTTGGCTACGG*	3	Chr08	2.9E+07	2.9E+07	intergenic:mRNA:MD08G122600-mRNA:MD08G122700
553rev	CCAGAACATGTTGGCTACAGG	CCGGAAGATGTTGGCTACGG*	3	Chr02	1.3E+07	1.3E+07	intergenic:mRNA:MD02G1159000-mRNA:MD02G1159100
553rev	CCAGAACATGTTGGCTACAGG	CCAGAACATGTTGGCTACGG*	3	Chr00	1.8E+07	1.8E+07	intergenic:mRNA:MD00G1087800-mRNA:MD00G1087900
553rev	CCAGAACATGTTGGCTACAGG	CCAGAACATGTTGGCTACGG*	4	Chr12	4504722	4504744	intergenic:mRNA:MD12G1040700-mRNA:MD12G1040800
553rev	CCAGAACATGTTGGCTACAGG	TCAGGAGATGTTGGCTACGG*	4	Chr16	1.3E+07	1.3E+07	intergenic:mRNA:MD16G1163700-mRNA:MD16G1163800
553rev	CCAGAACATGTTGGCTACAGG	GCAGGAGATGTTGGCTACGG*	4	Chr02	3.4E+07	3.4E+07	intergenic:mRNA:MD02G1280100-mRNA:MD02G1280200
553rev	CCAGAACATGTTGGCTACAGG	CTAGAACATGTTGGCTACGG*	4	Chr13	8378479	8378501	exon:mRNA:MD13G1115500
553rev	CCAGAACATGTTGGCTACAGG	CCAGGAGATGTTGGCTACGG*	4	Chr13	4.1E+07	4.1E+07	intergenic:mRNA:MD13G1282200-mRNA:MD13G1282300
553rev	CCAGAACATGTTGGCTACAGG	CCAGGAGATGTTGGCTACGG*	4	Chr13	4.1E+07	4.1E+07	intergenic:mRNA:MD13G1282200-mRNA:MD13G1282300
553rev	CCAGAACATGTTGGCTACAGG	CTAGAACATGTTGGCTACGG*	4	Chr09	1.5E+07	1.5E+07	intergenic:mRNA:MD09G1177600-mRNA:MD09G1177700
553rev	CCAGAACATGTTGGCTACAGG	CCGGAGATGTTGGCTACGG*	4	Chr02	3.4E+07	3.4E+07	intergenic:mRNA:MD02G1281900-mRNA:MD02G1282000
553rev	CCAGAACATGTTGGCTACAGG	CCGGAGATGTTGGCTACGG*	4	Chr03	2.9E+07	2.9E+07	intergenic:mRNA:MD03G1212300-mRNA:MD03G1212400
553rev	CCAGAACATGTTGGCTACAGG	CCGGAGATGTTGGCTACGG*	4	Chr05	3E+07	3E+07	intergenic:mRNA:MD05G1173100-mRNA:MD05G1173200
553rev	CCAGAACATGTTGGCTACAGG	CCGGAGATGTTGGCTACGG*	4	Chr04	6226469	6226491	intergenic:mRNA:MD04G1052900-mRNA:MD04G1053000
553rev	CCAGAACATGTTGGCTACAGG	CTAGAACATGTTGGCTACGG*	4	Chr04	1.9E+07	1.9E+07	intergenic:mRNA:MD04G1101300-mRNA:MD04G1101400
553rev	CCAGAACATGTTGGCTACAGG	CTAGAACATGTTGGCTACGG*	4	Chr01	2E+07	2E+07	intron:mRNA:MD01G1089100
553rev	CCAGAACATGTTGGCTACAGG	CCGGAGATGTTGGCTACGG*	4	Chr05	5468915	5468937	exon:mRNA:MD05G1034200
553rev	CCAGAACATGTTGGCTACAGG	CCGGAGATGTTGGCTACGG*	4	Chr12	8737425	8737447	intergenic:mRNA:MD12G1073200-mRNA:MD12G1073300
553rev	CCAGAACATGTTGGCTACAGG	CGAGAGATGTTGGCTACGG*	4	Chr12	4846962	4846984	intergenic:mRNA:MD12G1043000-mRNA:MD12G1043100
553rev	CCAGAACATGTTGGCTACAGG	CGAGCATGTTGGCTACGG*	4	Chr08	3607946	3607968	intergenic:mRNA:MD08G1047500-mRNA:MD08G1047600
553rev	CCAGAACATGTTGGCTACAGG	CCAGCATGTTGGCTACGG*	4	Chr16	4.1E+07	4.1E+07	intergenic:mRNA:MD16G1286700-mRNA:MD16G1286800
553rev	CCAGAACATGTTGGCTACAGG	CCAGAACATGTTGGCTACGG*	4	Chr11	3E+07	3E+07	intergenic:mRNA:MD11G1203300-mRNA:MD11G1203400
553rev	CCAGAACATGTTGGCTACAGG	CCAGAACATGTTGGCTACGG*	4	Chr03	217047	217069	exon:mRNA:MD03G1002900
553rev	CCAGAACATGTTGGCTACAGG	CCAGGAGATGTTGGCTACGG*	4	Chr00	4E+07	4E+07	intergenic:mRNA:MD00G1172500-mRNA:MD00G1172600
553rev	CCAGAACATGTTGGCTACAGG	CCAGGAGATGTTGGCTACGG*	4	Chr08	7158490	7158512	intergenic:mRNA:MD08G1086100-mRNA:MD08G1086200
553rev	CCAGAACATGTTGGCTACAGG	CCGGAAGATGTTGGCTACGG*	4	Chr02	1.3E+07	1.3E+07	intergenic:mRNA:MD02G1159000-mRNA:MD02G1159100
553rev	CCAGAACATGTTGGCTACAGG	CAAGAACATGTTGGCTACGG*	4	Chr12	1660997	1661019	intergenic:mRNA:MD12G1017000-mRNA:MD12G1017200
553rev	CCAGAACATGTTGGCTACAGG	GCAGAACATGTTGGCTACGG*	4	Chr11	3.6E+07	3.6E+07	intron:mRNA:MD11G1251000
553rev	CCAGAACATGTTGGCTACAGG	TCAGAACATGTTGGCTACGG*	4	Chr02	8729956	8729978	intergenic:mRNA:MD02G1108000-mRNA:MD02G1108200
553rev	CCAGAACATGTTGGCTACAGG	GCAGAACATGTTGGCTACGG*	4	Chr02	1.4E+07	1.4E+07	intergenic:mRNA:MD02G1166800-mRNA:MD02G1166900

407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAAAGATAATCCAACGG*..*..*	4	Chr11	3E+07	3E+07	intergenic:mrRNA:MD11G1207400-mrRNA:MD11G1207500
407forw	TGATGAGAAAGATAATCCAACGG	AGTTGAGCAGGATAATCCAAGGG	*..*..*.....	4	Chr06	489199	489221	exon:mrRNA:MD06G1004200
407forw	TGATGAGAAAGATAATCCAACGG	GAATAAGAATGATAATCCAAGG	*..*..*.....	4	Chr14	1.6E+07	1.6E+07	intron:mrRNA:MD14G1100600
407forw	TGATGAGAAAGATAATCCAACGG	TTATGTGATGGATAATCCAAGG	*..*..*.....	4	Chr13	2.8E+07	2.8E+07	intergenic:mrRNA:MD13G1254300-mrRNA:MD13G1254600
407forw	TGATGAGAAAGATAATCCAACGG	TGAAGATACAGATAAACCACAAAGG	*..*..*.....	4	Chr04	2594633	2594655	exon:mrRNA:MD04G1022000
407forw	TGATGAGAAAGATAATCCAACGG	CGACGAGAAAGTTAACCCAAAGG	*..*..*.....	4	Chr05	4.4E+07	4.4E+07	intergenic:mrRNA:MD00G1184500-mrRNA:MD00G1184900
407forw	TGATGAGAAAGATAATCCAACGG	AGATGAGTAAGATAATTAATATGG	*..*..*.....	4	Chr05	1.7E+07	1.7E+07	intergenic:mrRNA:MD05G1081300-mrRNA:MD05G1081400
407forw	TGATGAGAAAGATAATCCAACGG	TGATGTGATAGATACCAATATGG	*..*..*.....	4	Chr15	5.4E+07	5.4E+07	exon:mrRNA:MD15G1435300
407forw	TGATGAGAAAGATAATCCAACGG	TGATGTGATAGATACCAATATGG	*..*..*.....	4	Chr08	3.1E+07	3.1E+07	exon:mrRNA:MD08G1241300
407forw	TGATGAGAAAGATAATCCAACGG	CGACGAAAAGCTAACCCAAAGG	*..*..*.....	4	Chr14	1.8E+07	1.8E+07	intergenic:mrRNA:MD14G113600-mrRNA:MD14G113700
407forw	TGATGAGAAAGATAATCCAACGG	TGATGCACAAGATAATCCAAGG	*..*..*.....	4	Chr08	4089087	4089109	exon:mrRNA:MD08G1051900
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAAAGTAAGTCCATAGG*..*..*	4	Chr08	904535	904557	intergenic:mrRNA:MD08G1011800-mrRNA:MD08G1011900
407forw	TGATGAGAAAGATAATCCAACGG	TGCTGAGAAAAGATAATATATGG	*..*..*.....	4	Chr15	3.8E+07	3.8E+07	exon:mrRNA:MD15G1335700
407forw	TGATGAGAAAGATAATCCAACGG	CGATGAGAAAAGTTGACCCAAAGG	*..*..*.....	4	Chr05	1.7E+07	1.7E+07	intergenic:mrRNA:MD05G1082900-mrRNA:MD05G1083200
407forw	TGATGAGAAAGATAATCCAACGG	GGATGAGAAAAGATAAACCACAAAG	*..*..*.....	4	Chr08	1.2E+07	1.2E+07	intergenic:mrRNA:MD08G1129800-mrRNA:MD08G1129900
407forw	TGATGAGAAAGATAATCCAACGG	GGATGAGAAAAGATAAACCACAAAG	*..*..*.....	4	Chr07	1.8E+07	1.8E+07	intergenic:mrRNA:MD07G1126400-mrRNA:MD07G1126500
407forw	TGATGAGAAAGATAATCCAACGG	GGATGAGAAAAGATAAACCACAAAG	*..*..*.....	4	Chr16	1.8E+07	1.8E+07	intergenic:mrRNA:MD16G1197200-mrRNA:MD16G1197300
407forw	TGATGAGAAAGATAATCCAACGG	AGAAGAGAAAAGAGATCCCAATGG	*..*..*.....	4	Chr11	397150	397172	intergenic:mrRNA:MD11G1004300-mrRNA:MD11G1004500
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAAAGATAAACCACATGG	*..*..*.....	4	Chr07	2.8E+07	2.8E+07	exon:mrRNA:MD07G1206400
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAAAGATAAACCACATGG	*..*..*.....	4	Chr07	2.9E+07	2.9E+07	exon:mrRNA:MD07G1206700
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAAAGATAAACCACATGG	*..*..*.....	4	Chr07	2.8E+07	2.8E+07	exon:mrRNA:MD07G1206100
407forw	TGATGAGAAAGATAATCCAACGG	TGACGACAAAGCTAACCCAAAGG	*..*..*.....	4	Chr11	3.6E+07	3.6E+07	intergenic:mrRNA:MD11G1249300-mrRNA:MD11G1249400
407forw	TGATGAGAAAGATAATCCAACGG	TCATTAAAAGATAAAGCACAAGG	*..*..*.....	4	Chr09	8763626	8763648	intergenic:mrRNA:MD09G1114900-mrRNA:MD09G1115000
407forw	TGATGAGAAAGATAATCCAACGG	CGATGAGAAAAGCTGACCCAAAGG	*..*..*.....	4	Chr10	1.9E+07	1.9E+07	intergenic:mrRNA:MD10G116500-mrRNA:MD10G116600
407forw	TGATGAGAAAGATAATCCAACGG	CGATGAGAAAAGCTGACCCAAAGG	*..*..*.....	4	Chr00	3.5E+07	3.5E+07	intergenic:mrRNA:MD00G1159500-mrRNA:MD00G1159900
407forw	TGATGAGAAAGATAATCCAACGG	CGATGAGAAAAGCTGACCCAAAGG	*..*..*.....	4	Chr16	1.3E+07	1.3E+07	intergenic:mrRNA:MD16G1160100-mrRNA:MD16G1160200
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAGATAAACAACATGG	*..*..*.....	4	Chr02	2314772	2314794	intergenic:mrRNA:MD02G1029800-mrRNA:MD02G1029900
407forw	TGATGAGAAAGATAATCCAACGG	TGAACGAAAATATAACAAGG	*..*..*.....	4	Chr04	5213375	5213397	intron:mrRNA:MD04G1044700
407forw	TGATGAGAAAGATAATCCAACGG	TGACGAGAAAAGTTGACCCAAAGG	*..*..*.....	4	Chr06	1.1E+07	1.1E+07	exon:mrRNA:MD06G1062300-mrRNA:MD06G1062400
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGCAACATAATACAGAG	*..*..*.....	4	Chr07	713654	713676	exon:mrRNA:MD07G1007800
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAAAGATAAACCACAAAGG	*..*..*.....	4	Chr15	4.1E+07	4.1E+07	intergenic:mrRNA:MD15G1349400-mrRNA:MD15G1349500
407forw	TGATGAGAAAGATAATCCAACGG	TAAAGAGAAAAGTTAATCAATGG	*..*..*.....	4	Chr15	1.9E+07	1.9E+07	intergenic:mrRNA:MD15G1235700-ncRNA:MD15G1235800
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAAAGATAAATCAATGG	*..*..*.....	4	Chr04	2.4E+07	2.4E+07	intergenic:mrRNA:MD04G1148000-mrRNA:MD04G1148100
407forw	TGATGAGAAAGATAATCCAACGG	AGACGAGAAAAGATAATCGAAAGG	*..*..*.....	4	Chr14	1.9E+07	1.9E+07	intergenic:mrRNA:MD14G1117300-mrRNA:MD14G1117700
407forw	TGATGAGAAAGATAATCCAACGG	TGTTGATAAGCTAATCAATGG	*..*..*.....	4	Chr15	3.4E+07	3.4E+07	intron:mrRNA:MD15G1321300
407forw	TGATGAGAAAGATAATCCAACGG	TGATGCACAAGCTGACCCAAAGG	*..*..*.....	4	Chr00	4.6E+07	4.6E+07	intergenic:mrRNA:MD00G1191000-ncRNA:MD00G1191100
407forw	TGATGAGAAAGATAATCCAACGG	TGAAGAAAAGATATCAAAATGG	*..*..*.....	4	Chr16	1.5E+07	1.5E+07	intergenic:mrRNA:MD16G1176900-mrRNA:MD16G1177000
407forw	TGATGAGAAAGATAATCCAACGG	TGACGAGAAAAGCTGACCCAAAGG	*..*..*.....	4	Chr02	2.4E+07	2.4E+07	intergenic:mrRNA:MD02G1213900-mrRNA:MD02G1214100
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAAAAGCTGACCTAAAGG	*..*..*.....	4	Chr00	8954776	8954798	intergenic:mrRNA:MD00G1047600-mrRNA:MD00G1047900
407forw	TGATGAGAAAGATAATCCAACGG	TGAGGAGATGATATTCATATGG	*..*..*.....	4	Chr03	1.8E+07	1.8E+07	intron:mrRNA:MD03G1156300
407forw	TGATGAGAAAGATAATCCAACGG	AAATGAGAAAATTAATGCAATGG	*..*..*.....	4	Chr17	1.8E+07	1.8E+07	intergenic:mrRNA:MD17G1172000-mrRNA:MD17G1172100
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAAAAGCTGACCCAAAGG	*..*..*.....	4	Chr11	3.4E+07	3.4E+07	intergenic:mrRNA:MD11G1232400-mrRNA:MD11G1232500
407forw	TGATGAGAAAGATAATCCAACGG	TGATGAGAAAATTAATCGTTGGG	*..*..*.....	4	Chr04	3.2E+07	3.2E+07	intergenic:mrRNA:MD04G1242600-mrRNA:MD04G1242700
407forw	TGATGAGAAAGATAATCCAACGG	TGTTGAGAAAATTAATACAATGG	*..*..*.....	4	Chr10	1.9E+07	1.9E+07	intergenic:mrRNA:MD10G1114700-mrRNA:MD10G1114800
407forw	TGATGAGAAAGATAATCCAACGG	TGTTGAGAAAATTAATACAATGG	*..*..*.....	4	Chr11	3.4E+07	3.4E+07	intergenic:mrRNA:MD11G1232900-mrRNA:MD11G1233100
407forw	TGATGAGAAAGATAATCCAACGG	TGTTGAGAAAATTAATACAATGG	*..*..*.....	4	Chr13	1.5E+07	1.5E+07	intergenic:mrRNA:MD13G1181700-mrRNA:MD13G1181800
407forw	TGATGAGAAAGATAATCCAACGG	TGTTGAGAAAATTAATACAATGG	*..*..*.....	4	Chr10	1E+07	1E+07	intergenic:mrRNA:MD10G1073200-mrRNA:MD10G1073300
407forw	TGATGAGAAAGATAATCCAACGG	TGTTGAGAAAATTAATACAATGG	*..*..*.....	4	Chr01	1.1E+07	1.1E+07	intergenic:mrRNA:MD01G1032000-mrRNA:MD01G1032200


407forw	TGATGAGAAAAGATAAATCCAACGG	TGTTGAGAAAATTTAATACAATGG	*.....**.....*	4	Chr08	1.3E+07	1.3E+07	intergenic:mrna:MD08G1139200-mrna:MD08G1139500
407forw	TGATGAGAAAAGATAAATCCAACGG	TGTTGAGAAAATTTAATACAATGG	*.....**.....*	4	Chr06	2559996	2560018	intergenic:mrna:MD06G1020600-mrna:MD06G1020700
407forw	TGATGAGAAAAGATAAATCCAACGG	TGTTGAGAAAATTTAATACAATGG	*.....**.....*	4	Chr11	3.1E+07	3.1E+07	intergenic:mrna:MD11G1213000-mrna:MD11G1213100
407forw	TGATGAGAAAAGATAAATCCAACGG	TGTTGAGAAAATTTAATACAATGG	*.....**.....*	4	Chr11	3.5E+07	3.5E+07	intergenic:mrna:MD11G1241100-mrna:MD11G1241400
407forw	TGATGAGAAAAGATAAATCCAACGG	TGTTGAGAAAATTTAATACAATGG	*.....**.....*	4	Chr16	3.2E+07	3.2E+07	intergenic:mrna:MD16G1259500-mrna:MD16G1260200
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATCAGAAATATAAGCCGAAGG	*.....**.....*	4	Chr08	5162660	5162682	intergenic:mrna:MD08G1065000-mrna:MD08G1065100
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATGAGATAGATAAAGAAAAGG	*.....**.....*	4	Chr07	2.5E+07	2.5E+07	intergenic:mrna:MD07G1174700-mrna:MD07G1174800
407forw	TGATGAGAAAAGATAAATCCAACGG	TTATGAGGAAGCTACTCCAATGG	*.....**.....*	4	Chr02	1E+07	1E+07	exon:mrna:MD02G1127200
407forw	TGATGAGAAAAGATAAATCCAACGG	TGTTTTGAAAAGATCAATCCAAGGG	*.....**.....*	4	Chr08	1.1E+07	1.1E+07	intergenic:mrna:MD08G1116100-mrna:MD08G1116200
407forw	TGATGAGAAAAGATAAATCCAACGG	TGAGGAGAAAAGTTATTTCCAGAGG	*.....**.....*	4	Chr01	3.3E+07	3.3E+07	exon:mrna:MD01G1238900
407forw	TGATGAGAAAAGATAAATCCAACGG	TGAGGAGAAAAGTTATTTCCAGAGG	*.....**.....*	4	Chr07	3.7E+07	3.7E+07	exon:mrna:MD07G1310800
407forw	TGATGAGAAAAGATAAATCCAACGG	AGATGAGAGAGAAAATGCCAAGGG	*.....**.....*	4	Chr11	3.7E+07	3.7E+07	intron:mrna:MD11G1259100
407forw	TGATGAGAAAAGATAAATCCAACGG	TCATGAGACAGATAATTTGAATGG	*.....**.....*	4	Chr13	4491229	4491251	intergenic:mrna:MD13G1065000-mrna:MD13G1065100
407forw	TGATGAGAAAAGATAAATCCAACGG	TGAGGATAAAGAGAAATCCCAAGG	*.....**.....*	4	Chr00	1.3E+07	1.3E+07	intergenic:mrna:MD00G1067300-mrna:MD00G1067400
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATGCGAAGATATTTCAAGGG	*.....**.....*	4	Chr11	4.1E+07	4.1E+07	exon:mrna:MD11G1296700
407forw	TGATGAGAAAAGATAAATCCAACGG	TGTTGAGAAAATTTAATGCAAGGG	*.....**.....*	4	Chr14	355386	355408	intron:mrna:MD14G1002200
407forw	TGATGAGAAAAGATAAATCCAACGG	TGAAGAGAAAAGATAAAGAAAATGG	*.....**.....*	4	Chr10	3.6E+07	3.6E+07	exon:mrna:MD10G1261500
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATGAGAAAAGATAAAGAGAGGG	*.....**.....*	4	Chr11	4E+07	4E+07	intergenic:mrna:MD11G1278100-mrna:MD11G1278400
407forw	TGATGAGAAAAGATAAATCCAACGG	TGAGGAGAAAATCTGCAAAAGG	*.....**.....*	4	Chr02	2.5E+07	2.5E+07	exon:mrna:MD02G1217000
407forw	TGATGAGAAAAGATAAATCCAACGG	TGAGGAGAAAATCTGCAAAAGG	*.....**.....*	4	Chr02	3.2E+07	3.2E+07	exon:mrna:MD02G1263900
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATGAGAAAAGATAAAGAGAGGG	*.....**.....*	4	Chr16	1E+07	1E+07	exon:mrna:MD16G1132300
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATGAGAAAAGATAAAGAGAGGG	*.....**.....*	4	Chr05	9335527	9335549	intergenic:mrna:MD05G1054300-mrna:MD05G1054400
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATGAGAAAATTTATTCGAATGG	*.....**.....*	4	Chr01	3E+07	3E+07	exon:mrna:MD01G1205300
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATGAGAAAATTTATTCGAATGG	*.....**.....*	4	Chr07	3.4E+07	3.4E+07	intergenic:mrna:MD07G1274400-mrna:MD07G1274500
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATGAGAAAATTTATTCGAATGG	*.....**.....*	4	Chr07	3.4E+07	3.4E+07	exon:mrna:MD07G1273900
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATGAGAAAAGATAAATGGAATGG	*.....**.....*	4	Chr08	3.1E+07	3.1E+07	exon:mrna:MD08G1238500
407forw	TGATGAGAAAAGATAAATCCAACGG	TGATGAGCAAGAAAATGGAAGGG	*.....**.....*	4	Chr15	1.5E+07	1.5E+07	exon:mrna:MD15G1194600

Annexe 9: Protocole pour réaliser une mutation sur un plasmide par amplification complète

	Institut de Recherche en Horticulture et Semences Centre INRA Angers-Nantes 42 rue Georges Morel, 49071 Beaucouzé	Document number: Date of creation: December 12, 2019 Last revision: December 12, 2019
	Document: Homemade Site Directed Mutagenesis of Whole plasmids – Valema/Respom	

Purpose of the document: Create site directed mutation in plasmid in one step

Table of Creation

Created by (Name/Title)	Verified by (Name/Title)	Authorized by (Name/Title)
Adrien Perrin - PhD Date: December 12, 2019		
Signature 	Signature	Signature

Reference: Laible and Boonrod 2009 (NCBI)

Safety Precautions

To manipulate bacteria under safety fume hood

Procedures

3.1 Material

3.1.1 Consumables

Regular lab consumable
 Proofreading polymerase as Q5 high fidelity polymerase from NEB
 Top Ten® chemiocompetent cell (thermofisher)
 Degenerated primer with selected mutation

3.1.2 Solutions

Enzyme : DpnI methylation sensitive enzyme
 BclI : or your favorite enzyme to cleave plasmid and check DpnI result (indirect validation).

3.1.3 Equipment

Thermocycler (for a day or overnight)

3.2 Preparation

3.2.1 Long range PCR

Prepare the following mix for 1 reaction in 50µL

Reagent	Quantity
H2O	QSP: 50µL
Buffer (10X)	10µL
dNTP (10mM)	1µL
Primer 1 (10M)	2.5µL
Primer 2 (10M)	2.5µL
Q5 polymerase	1µL
DNA	1 to 50ng

And use the following programme

Step	Temperature (°C)	Time
1	98	30"
2	98	10"
3	60	30"
4	72	40"/kb +2'/10kb
5	Go to step 2 30 times	
6	72	40"/kb +2'/10kb +2'
7	4	∞

3.2.2 Remove remaining native plasmid

Perform digestion with methylation sensitive enzyme (that cleaving DNA only if DNA is methylated)
 Use enzyme at recommended concentration with respect to use DNA quantity in PCR.
 Incubate 1 hour at 37°C
 Then inactivate enzyme 20' at 80°C

3.2.2 Checking of DpnI activity


On aliquot of not digested plasmid and aliquot of digested plasmid perform digestion with BclI and check this digestion on agarose gel.
 Do not forget to keep one aliquot digest by DpnI but not by BclI to following step

3.2.2 Transformation

Use the digested solution of mutated plasmid by DpnI to perform transformation. Use commercial chemo-competent cell or electrocompetent cell to maximize transformation rate. Grow the transformed cell on selective medium.


Version 01	IRHS – Valema / ResPom	Page 1 of 1
------------	------------------------	-------------

Annexe 10: Protocole pour réaliser deux mutations sur un plasmide via un golden gate

	Institut de Recherche en Horticulture et Semences Centre INRA Angers-Nantes 42 rue Georges Morel, 49071 Beaucouzé	<i>Document number:</i> Date of creation: December 12, 2019 Last revision: December 12, 2019
	<i>Document:</i> To obtain deadCas9 plasmid from plasmid containing Cas9 plasmids – Valema/Respom	

Purpose of the document: Create two mutation in Cas9 to inactivate it on domesticated plasmid. You will create mutation at two different sites. Here we propose to amplify Cas9 in two pieces, creating a mutation on each piece, then ligate this to piece to recover the complete sequence (with the 2 mutation) and clone it back in your domesticated plasmid. This is usable to create mutations on other specific gene in your favorite plasmid.

Table of Creation

Created by (Name/Title)	Verified by (Name/Title)	Authorized by (Name/Title)
Adrien Perrin - PhD Date: December 12, 2019		
Signature 	Signature	Signature

Reference: Golden Gate cloning: (Engler et al. 2008)

Safety Precautions

To manipulate bacteria under safety fume hood

Procedures

3.1 Material

3.1.1 Consumables

- Regular lab consumable
- Proofreading polymerase as Q5 high fidelity polymerase from NEB
- Top Ten® chemiocompetent cell (thermofisher)
- Degenerated primer with selected mutation
- pJet cloning kit
- T4 DNA ligase

3.1.2 Solutions

Enzyme : your favorite enzyme

3.1.3 Equipment

Thermocycler (for a day or overnight)

3.2 Preparation

3.2.1 Long range PCR

Using 2 primer couple. The mutation that you want to create must to be in the middle of the primer. Perform PCR with Q5 high fidelity and proofreading polymerase (Classic polymerase add polyA in 3' that is not needed here).

Prepare the following mix for 1 reaction in 50µL

Reagent	Quantity
H2O	QSP: 50µL
Buffer (10X)	10µL
dNTP (10mM)	1µL
Primer 1 (10M)	2.5µL
Primer 2 (10M)	2.5µL
Q5 polymerase	1µL
DNA	1 to 50ng

And use the following programme

Step	Temperature (°C)	Time
1	98	30"
2	98	10"
3	60	30"
4	72	40"/kb +2'/10kb
5	Go to step 2 30 times	
6	72	40"/kb +2'/10kb +2'
7	4	∞


You obtained 2 amplicons from this PCR each with needed mutation

3.2.2 PCR2: Amplification and addition restriction site

This step increase amount of DNA for following step.

Use new primer couple to add your favorite enzyme. We recommended you to use GoldenGate cloning with addition of BsaI site to ligate the to piece of the Cas9, please refere to Engler et al (2008) . Carefully check the presence of extra BsaI site on the sequence.

Version 01	IRHS – Valema / ResPom	Page 1 of 2
------------	------------------------	-------------

	Institut de Recherche en Horticulture et Semences Centre INRA Angers-Nantes 42 rue Georges Morel, 49071 Beaucouzé	<i>Document number:</i> <i>Date of creation:</i> December 12, 2019
	<i>Document:</i> To obtain deadCas9 plasmid from plasmid containing Cas9 plasmids – Valema/Respom	<i>Last revision:</i> December 12, 2019

3.2.3 Golden Gate cloning

Prepare the following mix

Reagent	Quantity (µL)
H2O (Qsp 20µL)	8
T4 Buffer	2
Amplicon 1	4 (50ng/µL)
Amplicon 2	4 (50ng/µL)
T4 DNA ligase	1,25
BsaI	0,50

And use the following programme

Step	Temperature (°C)	Time
1	37	2'
2	16	5'
3	Go to step 1, 50 times	
4	50	5'
5	80	5'
6	4	∞

Check on agarose gel the correct ligation
 Then purify the remaining part and quantify it

3.2.4 Cloning in pJet

DNA quantity recommended 200ng. Use a ratio of 3:1 in favor of the insert (ligation product)
 Perform pJet ligation as following

Reagent	Quantity (µL)
Buffer	10
Ligation product	15
pJET1.2/blunt cloning vector (50ng/µL)	1
T4 DNA ligase	1

Incubate at room temperature (22°C) for 2-3h
 Then transform, select and verified it (sequencing) before to continue to next step

3.2.5 enzymatic restriction:

pJET deadCas9: here we use the AscI enzyme to open plasmid. Then we purified it.
 pDE-cas9 (or your domesticated plasmid): because of additional modification we open the plasmid with AscI and MfeI. Then we dephosphorylated it with Antarctic Phosphatase following recommended procedure. Then we purified the open plasmid to remove enzyme (MfeI is not heat inactivable)

3.2.6 Final ligation:


Perform T4 ligation as recommend. We recommended to you to try multi ratio of insert and plasmid ligation. Do not hesitate to increase ligation time to 48h at 4°C if large plasmid. We recommended to use electrocompetent cell.

Annexe 11: Protocole de pour produire des bactéries électro-compétentes à partir de bactéries thermo-compétentes ou portant la résistance au gène de sélection ccdb.

	Institut de Recherche en Horticulture et Semences Centre INRA Angers-Nantes 42 rue Georges Morel, 49071 Beaucouzé	Document number: Date of creation: December 12, 2019
	Document: Electrocompetent cell production and electroporation procedure Valema/Respom	Last revision: December 12, 2019

Purpose of the document: Produce large amount of home made electrocompetent cell

Table of Creation

Created by (Name/Title)	Verified by (Name/Title)	Authorized by (Name/Title)
Adrien Perrin - PhD Date: December 12, 2019		
Signature 	Signature	Signature

Reference: personal communication

Safety Precautions

To manipulate bacteria under safety fume hood

Procedures:

Electro-competent cells

Work in a cold environment

Day before:

- prepare 1 liter **LB without NaCl**, 1.5 liter H₂O, 25 ml 20% glycerol, ± 100x eppendorf tube, 10x 5 ml pipette point, 2x 500 ml centrifuge tube and 2x 10ml whitecap.
- make a 15 ml overnight culture with competent cells

First thing in the morning: inoculate the prewarmed LB with 1/100 volume. **2*200ml** bacteria. Store the water and glycerol at 4°C.

Grow the bacteria for approximately 3 hours until the OD₆₀₀ = 0.6 – 0.9

Chill the erlenmeyer 15 – 30 min. on ice

Centrifuge the culture 15 min. at 4000 rpm, 4°C. 8*50ml

Resuspend the pellet in **H₂O** and q again. **50ml H₂O/falcon**

Resuspend the pellet in **H₂O** and q again. **25ml H₂O/falcon**

Resuspend the pellet in **20 ml 20% glycerol** and q again. **5ml H₂O/2 falcons**

Resuspend the pellet in 2 - 3 ml 20% glycerol. **2ml/4tubes**

Divide into 50 ml aliquots.

Immediately freeze samples in liquid N₂, store competent cells at -80°C.

Testing competence:

Dilute pSK+ bluescript vector to 1 ng / ml. Electroporate (as usual) 50 ml cells with 1 ml plasmid and plate out 100 ml of recuperated bacteria. Plate out 100 ml of 10x and 100x dilution as well.

Count the colonies and calculate the equivalent of 1 mg plasmid. This is the competence in colony forming units (cfu).

Electroporation

Add 5 ml ligated plasmid to 50 ml competent *E. coli* cells

Transport the mix to a small electroporationcuvet (1 mm, Eurogentech)

Apply shock of 1.5 kV

Quickly add 500 ml LB-medium (no antibiotic!)

Transport mix to a new eppendorf tube

Let the transformed cells recover at 37° for 45 min.

Plate out different quantities (close to flame, e.g. 10 & 75 ml or more)

Incubate overnight at 37° upside down

Store further at 4°

Use a sterile wooden tip to pick useful colonies

and brush them on a new plate for screening

Version 01	IRHS – Valema / ResPom	Page 1 of 2
------------	------------------------	-------------

Annexe 12: séquence de la deadCas9

> DeadCas9 attendue

> SNP trouvés au séquençage

```
TTCGGCGCGC CATC GATAAGAAGTACTCTATCGG ACTCG G ATATCGGAACTAACTCTGTGGGATGGGCTGTGATCACCGATGAGTACA
-----
AGGTGCCATCTAAGAAGTTCAAGGTTCTCGGAAACACCGATAGGCACTCTATCAAGAAAAACCTTATCGGTGCTCTCCTCTTCGATT
-----
CTGGTGAAACTGCTGAGGCTACCAGACTCAAGAGAACCCTAGAGAAGGTACACCAGAAGAAAGAACAGGATCTGCTACCTCCAAG
-----
AGATCTTCTCTAACGAGATGGCTAAAGTGGATGATTCAATCTTCCACAGGCTCGAAGAGTCATTCCTCGTGAAGAAGATAAGAAGC
-----
ACGAGAGGCACCTATCTTCGAAACATCGTTGATGAGGTGGCATAACCACGAGAAGTACCCTACTATCTACCACCTCAGAAAGAAGC
-----
TCGTTGATTCTACTGATAAGGCTGATCTCAGGCTCATCTACCTCGCTCTCGCTCACATGATCAAGTTCAGAGGACACTTCCTCATCG
-----
AGGGTGATCTCAACCCTGATAACTCTGATGTGGATAAGTTGTTTCATCCAGCTCGTGCAGACCTACAACCAGCTTTTCGAAGAGAACC
-----
CTATCAACGCTTCAGGTGTGGATGCTAAGGCTATCCTCTCTGTAGGCTCTCTAAGTCAAGAAGGCTTGAGAACCTCATTGCTCAGC
-----
TCCCTGGTGAGAAGAAGAACGGACTTTTCGAAACTTGATCGCTCTCTCTCTCGGACTCACCCCTAACTTCAAGTCTAACTTCGATC
-----
TCGCTGAGGATGCAAAGCTCCAGCTCTCAAAGGATACCTACGATGATGATCTCGATAACCTCCTCGCTCAGATCGGAGATCAGTACG
-----
CTGATTTGTTCTCCTCGCTGCTAAGAACCTCTCTGATGCTATCCTCCTCAGTGATATCCTCAGAGTGAACACCGAGATCACCAAGGCTC
-----
CACTCTCAGCTTCTATGATCAAGAGATACGATGAGCACCACCAGGATCTCACACTTCTCAAGGCTCTTGTTAGACAGCAGCTCCCAG
-----
AGAAGTACAAAGAGATTTTCTTCGATCAGTCTAAGAACGGATACGCTGGTTAC ATCGATGGTGGTGCATCTCA GAAGAGTTCTACA
-----
AGTTCATCAAGCCTATCCTCGAGAAGATGGATGGAACCGA GGAACTCCTCGTGAAGCTCAA TAGAGAGGATCTTCTCAGAAAGCAGA
-----
GGACCTTCGATAACGGATCTATCCCTCATCAGATCCACCTCGGAGAGTTGCACGCTATCCTTAGAAGGCAAGAGGATTTCTACCCAT
-----
TCCTCAAGGATAACAGGAAAAGATTGAGAAGATTCTCACCTTCAGAATCCCTTACTACGTGGGACCTCTCGCTAGAGGAAACTCAT
-----
AGATTCGCTTGATGACCAGAAAGTCTGAGGAAACCATCACCCCT TGGAACTTCGAAGAGGTGGTGGATAAGGGTGCTAGTGCTCAG
-----
A
-----
TCTTTCATCGAGAGGATGACCAACTTCGATAAGAACCTTCCAAACGAGAAGGTGCTCCCTAAGCACTCTTTGCTCTACGAGTACTTC
-----
ACCGTGTACAACGAGTTGACCAAGGTTAAGTACGTGACCGAGGGAATGAGGAAGCCTGCTTTTTTGTCAGGTGAGCAAAAAGAAGGCT
-----
ATCGTTGATCTCTTGTTCAAGACCAACAGAAAGGTGACCGTGAAGCAGCTCAAAGAGGATTACTTCAAGAAAATCGAGTGCTTCGAT
-----
TCAGTTGAGATTTCTGGTGTGAGGATAGGTTCAACGCATCTCTCGGAACCTACCACGATCTCCTCAAGATCAT T AAGGATAAGGAT
-----
G
```

TTCTTGGATAACGAGGAAAACGAGGATATCTTGGAGGATATCGTTCTTACCCTCACCCCTCTTTGAAGATAGAGAGATGATTGAAGAA

 AGGCTCAAGACCTACGCTCATCTCTTCGATGATAAGGTGATGAAGCAGTTGAAGAGAAGAAGATACTGGTTGGGGAAGGCTCTCA

 AAGAAAGCTCATTAACGGAATCAGGGATAAGCAGTCTGGAAAGACAATCCTTGATTTCTCAAGTCTGATGGATTTCGTAACAGAAA

 CTTCATGCAGCTCATCCACGATGATTCTCTCACCTTTAAAGAGGATATCCAGAAGGCTCAGGTTTCAGGACAGGGTGATAGTCTCCA

 TGAGCATATCGCTAACCTCGCTGGATCTCTGCAATCAAGAAGGGAATCCTCCAGACTGTGTGAAGGTTGTGGAATGAGTTGGTGAAG

 CTGATGGCAAGGCATAAGCCTGAGAACATCGTGATCGAAATGGCTAGAGAGAACCAGACCACTCAGAAGGGACAGAAGAACTCTAGG

 GAAAGGATGAAGAGGATCGAGGAAGGTATCAAAGAGCTTGGATCTCAGATCCTCAAAGAGCACCCGTGTTGAGAACACTCAGCTCCAG

 AATGAGAAGCTCTACCTCTACTACCTCCAGAACGGAAGGGATATGTATGTGGATCAAGAGTTGGATATCAACAGGCTCTCTGATTAC

 GATGTTGATCTATCGTGCCACAGTCATTCTTGAAGGATGATTCTATCGATAACAAGGTGCTCACCAGGTCTGATAAGAACAGGGGT

 AAGAGTGATAACGTGCCAAGTGAAGAGGTTGTGAAGAAAATGAAGAACTATTGGAGGCAGCTCCTCAACGTAAGCTCATCACTCAG

 AGAAAGTTCGATAAAGTACTAAGGCTGAGAGGGGAGGACTCTCTGAATTGGATAAAGCAGGATTCATCAAGAGGCAGCTTGTGGAA

 ACCAGGCAGATCACTAAGCACGTTGCACAGATCCTCGATTCTAGGATGAACAACAAGTACGATGAGAACGATAAGTTGATCAGGGAA

 GTGAAGGTTATCACCCCTCAAGTCAAAGCTCGTGTCTGATTTTCAGAAAGGATTTCCAATTTACAAGGTGAGGGAATCAACAACCTAC

 CACCACGCTCAGATGCTTACCTAACGCTGTTGTTGGAACCGCTCTCATCAAGAAGTATCCTAAGCTCGAGTCAGAGTTCGTGTAC

 GGTGATTACAAGGTGTACGATGTGAGGAAGATGATCGCTAAGTCTGAGCAAGAGATCGGAAAGGCTACCGCTAAGTATTTCTTCTAC

 ACTCTAACATCATGAATTTCTTCAAGACCGAGATTACCCTCGCTAACGGTGAGATCAGAAAGAGGCCACTCATCGAGACAAACGGTG

 AAACAGGTGAGATCGTGTGGGATAAGGGAAGGATTTTCGCTACCGTTAGAAAGGTGCTCTCTAATGCCACAGGTGAACATCGTTAAGA

 AAACCGAGGTGCAGACCGGTGGATTCTCTAAAGAGTCTATCCTCCCTAAGAGGAACTCTGATAAGCTCATTGCTAGGAAGAAGGATT

 GGGACCCTAAGAAAATACGGTGGTTTCGATTCTCCTACCGTGGCTTACTCTGTTCTCGTTGTGGCTAAGGTTGAGAAGGGAAAGAGTA

 AGAAGCTCAAGTCTGTTAAGGAACTTCTCGGAATCACTATCATGGAAAGGTCATCTTTTCGAGAAGAACCAATCGATTTCTCGAGG

 CTAAGGGATACAAGAGGTTAAGAAGGATCTCATCATCAAGCTCCAAAGTACTCACTCTTCGAACTCGAGAACGGTAGAAAGAGGA

 TGCTCGCTTCTGCTGGTGAGCTTCAAAGGGAAACGAGCTTGCTCTCCATCTAAGTACGTTAACTTTCTTTACCTCGCTTCTCACT

```
ACGAGAAGTTGAAGGGATCTCCAGAAGATAACGAGCAGAAGCAACTTTTCGTTGAGCAGCACAAAGCACTACTTGGATGAGATCATCG
-----
AGCAGATCTCTGAGTTCTCTAAAAGGGTGATCCTCGCTGATGCAAACCTCGATAAGGTGTTGTCTGCTTACAACAAGCACAGAGATA
-----
AGCCTATCAGGGAACAGGCAGAGAACATCATCCATCTCTTACCCTTACCAACCTCGGTGCTCCTGCTGCTTTCAAGTACTTCGATA
-----
CAACCATCGATAGGAAGAGATACACCTCTACCAAAGAAGTGCTCGATGCTACCCTCATCCATCAGTCTATCACTGGACTCTACGAGA
-----
CTAGGATCGATCTCTCACAGCTCGGTGGTGATTCAAGGGCTGATCCTAAGAAGAAGAGGAAGGTTGAGGGCGCGCCGAG
-----
```

Annexe 13 : Schéma du plasmide pDE-Cas9-kr avec les sites de restrictions problématiques indiqués en rouge. Les sites attendus sont indiqués en vert.

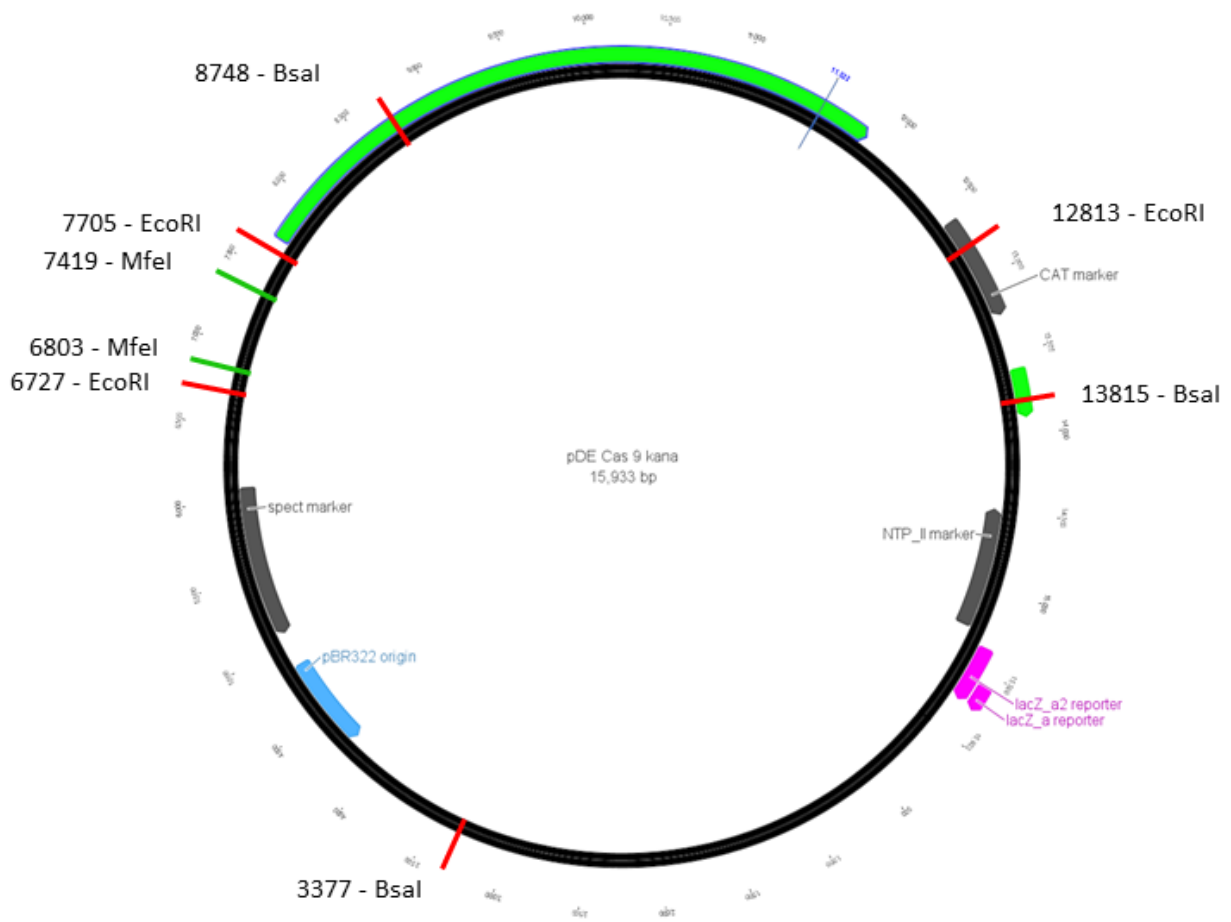


Table des matières

ABREVIATION	1
INTRODUCTION GENERALE	3
1. Malus domestica Borkh	5
1.1. Origine	5
1.2. Production mondiale et Française	5
1.3. Biologie	5
2. Multiplications sexuée et asexuée	7
2.1. La reproduction sexuée	7
2.1.1. Cas du pommier	7
2.2. La multiplication asexuée	9
2.2.1. Origine du greffage	9
2.2.2. Fonction et mise en place de la greffe	9
2.2.3. Epigénétique et greffage	10
2.2.4. Intérêt du greffage chez le pommier	11
3. La génétique chez les plantes	13
3.1. Le génome des plantes	13
3.1.1. Le génome du pommier	15
a) Les versions du génome de GDDH13	16
3.2. Transcription de l'ADN	17
3.2.1. Les ARNs messagers ou ARNs codants	17
3.2.2. Les ARNs non-codants	17
3.3. Traduction de l'ARNm	18
3.4. Régulation de l'expression des gènes	18
3.4.1. Régulation génétique	19
3.4.2. Régulation post-transcriptionnelle	19
4. Régulation épigénétique chez les plantes	21
4.1. Définition de l'épigénétique	21
4.2. Les histones	21
4.2.1. Définition	21
4.2.2. Modification des histones	21
a) L'acétylation de la queue histone	21
b) La méthylation de la queue histone	22
c) Acétylation et méthylation des histones chez le pommier	22
4.3. La méthylation de l'ADN	23
4.3.1. Définition	23
a) Chez le pommier	23
4.3.2. Rôle de la méthylation de l'ADN	24
a) Méthylation de l'ADN et éléments transposables	24
Méthylation des éléments transposable chez le pommier	25
Méthylation de l'ADN et régulation de la transcription des gènes	25
Méthylation et régulation de l'expression des gènes chez le pommier	26
4.3.3. Mécanisme de méthylation de l'ADN	27
a) La maintenance de la méthylation	27
b) La méthylation de novo	29
4.3.4. Méthylation de l'ADN et différences entre plantes pérennes et annuelles	29
4.3.5. Méthylation de l'ADN et amélioration des plantes	30
5. Transmission de l'information épigénétique	31
5.1. Multiplication asexuée	31
5.1.1. Cas du greffage	31
5.1.2. Cas de la multiplication in vitro	31
5.2. Multiplication sexuée	32
5.2.1. Gamétogénèse mâle	32
5.2.1. Gamétogénèse femelle	32
5.2.2. Fécondation	33
6. Objet de la thèse	35

6.1.	Contexte.....	35
6.2.	Objectif	35
CHAPITRE I : DIFFERENCES EPIGENETIQUES ENTRE UNE MULTIPLICATION SEXUEE ET UNE PROPAGATION ASEXUEE CHEZ LE POMMIER		37
1.	Introduction	39
2.	Matériels et méthodes spécifiques	41
2.1.	Matériel végétal.....	41
2.2.	Séquençage bisulfite	41
3.	Partie 1 – “Divergent DNA methylation signatures of juvenile seedlings, grafts and adult apple trees” (Perrin et al. 2020)	43
3.1.	Introduction à l'article	43
3.2.	Divergent DNA methylation signatures of juvenile seedlings, grafts and adult apple trees	43
3.3.	Résumé de l'article	64
4.	Partie 2 - Résultats complémentaires	65
4.1.	Matériels et méthodes	65
4.1.1.	Matériel végétal.....	65
4.1.2.	Fixation et coupes histologiques des échantillons YG et SD.	65
4.1.3.	Microarray et calcul des régions différenciellement méthylées.....	66
4.2.	Résultats	66
4.2.1.	Analyse de la troisième comparaison : plants greffés versus plants semés.....	66
a)	Phénotypage	66
	Phénotypage macroscopique.....	67
	Phénotypage microscopique	71
b)	Transcrits différenciellement exprimés.....	73
c)	Régions différenciellement méthylées	77
d)	Régions différenciellement méthylées en proximité des gènes	79
4.2.2.	Comparaison de deux méthodes de calculs utilisées pour l'analyse des DMRs	80
a)	Comparaison globale des deux méthodes de calculs	80
b)	Première méthode : Découpage en régions autour des gènes.....	82
c)	Seconde méthode : Sliding windows 200pb	85
4.3.	Discussion.....	86
4.3.1.	Effets de la transmission sexuée ou asexuée des marques épigénétiques	86
a)	Effets phénotypiques.....	86
	Choix du phénotype pour la publication	87
b)	Effet sur le niveau de transcription	88
c)	Régions différenciellement méthylées	88
d)	Est-ce que l'analyse de la comparaison YGvsSD corrobore les observations faites précédemment pour SDvsOG et OGvsYG ?	89
4.3.2.	Discussion des méthodes de calculs des DMRs utilisées et proposition d'autres méthodes de calculs	90
a)	Était-il judicieux de faire un calcul sur le génome complet plutôt que sur des régions spécifiques ?	90
b)	Pourquoi avoir choisi de faire les calculs avec la deuxième méthode plutôt que la première ? ...	91
c)	Quelles autres méthodes de calcul aurions-nous pu mettre en place et quelles seraient leurs avantages.....	92
	Méthode 3 : Calcul des DMRs en fonction du brin	94
	Méthode 4 : Calcul des DMRs en fonction et la structure des gènes	95
	Méthode 5 : prise en compte du profil de méthylation.....	96
	Méthode 6 : Analyse en fonction de la classification fonctionnel des gènes	97
d)	Quelle méthode choisir ?	97
4.3.3.	Critiques de différent points	98
a)	L'utilisation de la classification par code BIN : une bonne idée ?	98
b)	Discussion du plan expérimental	98
	Pertinence du prélèvement la première année	98
	Pertinence du modèle double haploïde.....	100
	Le double haploïde est un modèle d'étude simple	100
	Conséquence de l'autofécondation	100
5.	Conclusion	101
CHAPITRE II – EDITION DU GENOME ET DE L'EPIGENOME DU POMMIER.....		103

1.	Introduction	105
1.1.	<i>La technique CRISPR-Cas9</i>	105
1.1.1.	<i>Composition du système CRISPR-Cas9 et modalités de réparation des cassures double brin de l'ADN</i>	105
1.1.2.	<i>Evolution des applications de l'édition du génome</i>	107
1.1.3.	<i>Etat de l'art des technologies CRISPR-Cas9 chez le pommier</i>	108
1.2.	<i>Mutagenèse dirigée du gène MET1</i>	108
1.3.	<i>Modification de l'épigénome du promoteur du gène TFL1</i>	108
2.	Tentative d'extinction d'une voie de maintenance de la méthylation chez le pommier par édition des gènes codant pour la fonction MET1	111
2.1.	<i>Matériel et méthodes</i>	111
2.2.	<i>Résultats</i>	115
2.2.1.	<i>Recherche des orthologues de AtMet1 chez le pommier</i>	115
	a) V1.0 du génome de GDDH13	115
	b) V1.1 du génome de GDDH13	117
	c) Chez le génotype 'Gala'	118
2.2.2.	<i>Mise en place de la technique CRISPR-Cas9 pour réaliser l'extinction des gènes Md02MET1 et Md15MET1 du pommier</i>	119
	a) Sélection des crRNAs pour le génotype GDDH13	119
	b) Sélection des crRNAs pour le génotype 'Gala'	120
2.2.3.	<i>Résultats des transformations de vitroplants de pommiers</i>	120
2.2.4.	<i>Analyse moléculaire des lignées transformées</i>	122
2.2.5.	<i>Phénotypage en serre des lignées mutantes</i>	124
2.2.6.	<i>Caractérisation épigénétique des mutants</i>	126
	a) Epigénotypage	126
	b) Séquençage bisulfite du génome complet	129
2.3.	<i>Discussion</i>	129
2.3.1.	<i>Efficacité de transformation</i>	129
	a) Tentative de transformation du génotype GDDH13	129
2.3.2.	<i>Un taux important de profil d'édition homozygote</i>	130
2.3.3.	<i>La lignée 2P un hétérozygote pour toujours ?</i>	131
2.3.4.	<i>Les lignées 2C2 et 2C3 sont peut-être une seule et même lignée</i>	131
2.3.5.	<i>Séquençage bisulfite et test de digestion</i>	131
2.3.6.	<i>Phénotypes attendus pour des plantes dont les voies de maintien de la méthylation sont défectueuses</i>	132
3.	Edition de l'épigénome du pommier – modification de la méthylation d'un gène pour modifier son expression	135
3.1.	<i>Choix méthodologiques en vue de l'obtention d'une construction permettant la méthylation d'une séquence cible</i>	135
3.1.1	<i>Choix des crRNA ciblant le promoteur de MdTFL1</i>	135
3.1.2.	<i>Choix de la méthyltransférase et de son promoteur</i>	135
3.1.3	<i>Choix et obtention de la Cas9 inactivée</i>	137
	a) Premier protocole : mutation directement sur le plasmide pDeCas9-Kr (Figure 3.1-3 et Annexe 9)	139
	b) Second protocole : mutation par amplifications et clonages successifs à partir du plasmide pDeCas9-Kr (Figure 3.1-4, Annexe 10)	140
	c) Troisième protocole : utilisation d'une deadCas9 d'un plasmide existant (Figure 3.1-5)	142
3.1.2.	<i>Obtention de la portion HSP::DRM2 (Figure 3.1-1E)</i>	143
3.1.3.	<i>Obtention d'un plasmide contenant la construction HSP::DRM2::deadCas9 à partir, ou non, de pDe-deadCas9-Kr</i>	143
	a) Protocole d'obtention par Golden Gate (Figure 3.1-6)	143
	b) Protocole d'obtention par clonage conventionnel (Figure 3.1-6)	144
	c) Protocole d'obtention à l'aide d'un plasmide intermédiaire (Figure 3.1-7)	145
	d) Protocole d'insertion de la construction HSP::DRM2::deadCas9 synthétisée (Figure 3.1-8)	146
3.2.	<i>Résultats et discussion</i>	147
3.2.1.	<i>Obtention des crRNAs ciblant le promoteur de MdTFL1</i>	147
3.2.2.	<i>Obtention d'une deadCas9 fonctionnelle</i>	149
	a) Premier protocole : mutation directement sur le plasmide pDeCas9-Kr (Figure 3.1-3)	149
	b) Second protocole : mutation par amplifications et clonages successifs à partir du plasmide pDeCas9-Kr (Figure 3.1-4)	149
	c) Troisième protocole : Utilisation d'une deadCas9 d'un plasmide existant (Figure 3.1-5)	150
3.2.3.	<i>Echec de l'obtention de la construction HSP::DRM2 ::deadCas9 dans le plasmide final</i>	150

a)	Choix de la méthyltransférase DRM2	150
b)	Obtention de la construction HSP::DRM2 ::deadCas9 par Golden Gate (Figure 3.1-6)	150
c)	Obtention de la construction HSP::DRM2 ::deadCas9 par clonage conventionnel (Figure 3.1-6)	151
d)	Protocole d'obtention à l'aide d'un plasmide intermédiaire (Figure 3.1-7)	151
e)	Protocole d'insertion de la construction HSP::DRM2::deadCas9 synthétisée (Figure 3.1-8)	151
3.2.4.	<i>Proposition d'un protocole amélioré pour l'obtention de la construction HSP::DRM2 ::deadCas9 dans le plasmide final</i>	152
a)	Synthèse complète	152
b)	Protocole à mettre en place pour obtenir le plasmide pDE-HSP-DRM2-deadCas9-kr	152
4.	Conclusion – Chapitre II	155
	CONCLUSION GENERALE	157
	BIBLIOGRAPHIE	163
	ANNEXES	183
	Annexes à l'article (figures supplémentaires)	183
	Annexes Chapitre I	190
	Annexes Chapitre II	196
	TABLE DES ILLUSTRATIONS	219
	TABLE DES TABLEAUX	221

Table des illustrations

Figure 2.1-1 : Anatomie d'une fleur de pommier en A.....	8
Figure 2.2-1: modèle proposé par Berger et al 2018:	11
Figure 3.1-1 : structure et repliement de l'ADN (illustration par Georges Dolisi)	13
Figure 3.1-2: représentation de l'ADN et des étapes de transcription et traduction.	14
Figure 3.1-3 : diagramme de synténie du génome du pommier proposé par Daccord et al 2017	16
Figure 3.1-4 : Comparaison du réassemblage et de la ré-annotation des version V1.0 et V1.1 du génome de GDDH13.	17
Figure 3.3-1: La traduction	18
Figure 4.2-1: Illustration d'histones.	22
Figure 4.3-1: représentation de la méthylation d'une cytosine à l'échelle moléculaire	23
Figure 4.3-2: Modèle de régulation de l'expression de gènes d'A.thaliana via la méthylation de l'ADN	26
Figure 4.3-3: représentation des mécanismes de maintenance de la méthylation et de la voie de la RdDM	28
Figure 4.3-4: figure 2 de la revue de Law & Jacobsen, (2010)	29
Figure 5.2-1: Schéma de la reproduction sexuée chez A. thaliana issue de la revue	33
Figure 2.2-1 : schéma des étapes d'un séquençage bisulfite.....	42
Figure 3.2-1: Graphical abstract of (Perrin <i>et al.</i> , 2020) publication	43
Figure 4.1-1: illustration du prélèvement sur une feuille de pommier.	66
Figure 4.2-1 : comptage du nombre d'étages foliaires.	67
Figure 4.2-2: plan de prélèvement	68
Figure 4.2-3: photos macroscopiques de feuilles de pommier.	69
Figure 4.2-4: photos de feuilles de pommier réalisées à la loupe binoculaire	70
Figure 4.2-5: Exemples de coupes transversales de jeunes feuilles.	72
Figure 4.2-6: mesure de l'épaisseur du limbe (A.) et du diamètre de la nervure centrale (B.).	72
Figure 4.2-7 : DEGs communs aux trois comparaisons.	74
Figure 4.2-8 : classification BIN des DEGs spécifiques à chaque comparaison.	75
Figure 4.2-9 : Nombre, répartition et taux de méthylation des DMRs.	77
Figure 4.2-10: DMR entre les comparaisons et DEG DMR. A. A. DMRs communes tout contexte confondu. B. gène DMR et DEG DMR tout contexte confondu.	78
Figure 4.2-11 : schéma des différentes méthodes de calculs de DMRs à proximité d'un gène.	82
Figure 4.2-12 : répartition des DMRs par rapport aux gènes.	84
Figure 4.3-1 : Modèle général représentant les changements phénotypiques, transcriptomiques et épigénétiques ayant lieu lors d'une multiplication sexuée ou d'une propagation asexuée chez <i>Malus domestica</i>	90
Figure 4.3-2: autres méthodes de calcul envisageables	93
Figure 4.3-3 screenshot de JBrowse montrant un exemple de gap de méthylation.	94
Figure 4.3-4: quatrième méthode de calcul: représentation schématique de la méthode	95
Figure 4.3-5: cinquième méthode de calcul: représentation schématique de la méthode	96
Figure 4.3-6: Photos de pommier en serre et sous loupe binoculaire.	99
Figure 1.1-1 : Système CRISPR-Cas9.	107

Figure 1.1-2: Schéma de réparation d'une cassure double brin via le NHEJ ou la HDR.....	107
Figure 2.1-1: Motif de restriction de l'enzyme <i>BsaI</i> utilisé pour le Golden Gate et construction des gRNA.....	113
Figure 2.2-1: Structures géniques et protéiques des gènes codant pour la fonction MET1.....	116
Figure 2.2-2: Comparaison de la séquence des gènes codants pour la fonction MET1.....	117
Figure 2.2-3: Alignement entre les crRNA dessinés sur le génome de référence GDDH13.....	118
Figure 2.2-4 : Sélection des crRNAs sur le logiciel CRISPOR-Tefor.	120
Figure 2.2-5 : Vérification par PCR des lignées potentiellement mutantes pour le gène <i>MET1</i>	122
Figure 2.2-6: Analyse moléculaire des lignées mutantes..	124
Figure 2.2-7 : photographies de plants des lignées acclimatées en serre.	125
Figure 2.2-8 : Motif de clivage de <i>HpaII</i> et profil de méthylation du transposon HODOR.....	127
Figure 2.2-9 : Epigénotypage des lignées mutantes pour le gène MD15G1231100.....	128
Figure 3.1-1 : Description des différentes séquences d'ADN utilisées.....	136
Figure 3.1-2: Schéma des protocoles possibles pour l'obtention du plasmide	138
Figure 3.1-3: Schéma d'obtention des mutations nécessaires pour obtenir une deadCas9	140
Figure 3.1-4: Schéma d'un second protocole permettant d'obtenir le plasmide pDEadCas9.	141
Figure 3.1-5 : Schéma du protocole 3	142
Figure 3.1-6 : Schéma du protocole 4	144
Figure 3.1-7: Schéma du protocole 5	145
Figure 3.1-8 : Schéma du protocole 6	146
Figure 3.2-1 : Visualisation du gène <i>MdTFL1</i>	148
Figure 3.2-2: Proposition de protocole	153

Table des tableaux

Table 4.2-1 : Nombre de DEGs et DETEs par comparaison et en communs entre les comparaisons..	73
Table 4.2-2 : tableau récapitulatif des deux méthodes de calculs de DMRs utilisées.	81
Table 4.2-3 : nombre de DMRs calculé avec la première méthode de calculs.	83

RÉSUMÉ

Par son mode de propagation clonale (greffage), le pommier est un modèle intéressant pour les études épigénétiques. On observe en effet, une variabilité phénotypique au sein d'une même variété, suggérant une contribution de mécanismes épigénétiques. Au début de cette thèse les mécanismes de transmission de marqueurs épigénétiques, tel que la méthylation de l'ADN, étaient encore peu étudiés chez les plantes pérennes. Pour mieux comprendre comment les marques épigénétiques sont transmises et contribuent à des traits phénotypiques, nous avons défini deux axes de recherche principaux: (i) En utilisant la technologie CRISPR-Cas9 nous avons créé des lignées de pommier afin d'obtenir une diminution systématique de la méthylation de l'ADN. (ii) Afin d'évaluer la transmission des marques épigénétiques lors de la reproduction sexuée et asexuée, nous avons comparé les épigénomes d'arbres adultes, de greffons et de semis obtenus par autofécondation sur une lignée modèle double-haploïde. Les analyses à l'échelle de génomes complets ont montré une absence de modification globale au niveau de la méthylation de l'ADN entre un arbre adulte, un greffon et un semis de pommier. Cependant nous avons détecté des régions spécifiques présentant une différence de méthylation de l'ADN (DMRs). Fait intéressant, nous avons détecté plus de DMRs entre le semis et le greffon ou l'arbre qu'entre l'arbre et le greffon, ce qui suggère que l'arbre et le greffon sont plus proches au niveau épigénétique. Nos résultats de phénotypage, d'analyses transcriptomiques et de méthylome confirment la différence entre arbre adulte et semis. Mais également que les greffons sont à l'interface entre juvénilité et maturité.

mots-clés : *Malus domestica*, Plante pérenne, Méthylation de l'ADN, Multiplication sexuée, Propagation asexuée, CRISPR-Cas9, Transformation in vitro

ABSTRACT

Thanks to its clonal propagation (grafting), apple tree is an interesting model for epigenetic studies. Indeed, phenotypic variability can be observed within the same variety, suggesting a potential contribution of epigenetic mechanisms. At the beginning of this thesis, the mechanisms of transmission of epigenetic marks, such as DNA methylation, were still little studied in perennial plants. To better understand how epigenetic marks are transmitted and contribute to phenotypic traits, we have defined two main axes of research: (i) Using CRISPR-Cas9 technology we have created apple mutant lines in order to obtain a systematic decrease in DNA methylation. (ii) In order to assess the transmission of epigenetic marks during sexual and asexual reproduction, we compared the epigenomes of adult trees, grafts and seedlings obtained by self-fertilization on a model doubled-haploid line. Analyses at the whole epigenome level indicated that there is no overall change in DNA methylation between an adult tree, a grafted tree and an apple seedling. However, we were able to detect specific regions that are differentially methylated (DMRs). Interestingly, more DMRs were detected between seedlings and grafted tree or adult tree than between the tree and grafted tree, suggesting that the tree and grafted tree are closer epigenetically. Our phenotypic, transcriptomic and methylome analysis results confirm the difference between adult trees and seedlings. They also confirm that grafted tree is at the interface between juvenility and maturity.

keywords : *Malus domestica*, perennial plant, DNA methylation, Sexual multiplication, asexual propagation; CRISPR-Cas9, In vitro culture

Présidence de l'université
40 rue de rennes – BP 73532
49035 Angers cedex

Tél. 02 41 96 23 23 | Fax 02 41 96 23 00



ENGAGEMENT DE NON PLAGIAT

Je, soussigné Perrin Adrien déclare être pleinement conscient que le plagiat de documents ou d'une partie d'un document publiée sur toutes formes de support, y compris l'internet, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée.

En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour écrire ce rapport ou mémoire.

signé par l'étudiant(e) le **31/ 03 / 2020**

Perrin

Transmission des marques épigénétiques lors de la multiplication sexuée et de la propagation asexuée chez le pommier (*Malus domestica*)

Mots clés : *Malus domestica*, Plante pérenne , Méthylation de l'ADN, Multiplication sexuée , Propagation asexuée, CRISPR-Cas9 , Transformation *in vitro*

Résumé : Par son mode de propagation clonale (greffage), le pommier est un modèle intéressant pour les études épigénétiques. On observe en effet, une variabilité phénotypique au sein d'une même variété, suggérant une contribution de mécanismes épigénétiques. Au début de cette thèse les mécanismes de transmission de marqueurs épigénétiques, tel que la méthylation de l'ADN, étaient encore peu étudiés chez les plantes pérennes. Pour mieux comprendre comment les marques épigénétiques sont transmises et contribuent à des traits phénotypiques, nous avons défini deux axes de recherche principaux: (i) En utilisant la technologie CRISPR-Cas9 nous avons créé des lignées de pommier afin d'obtenir une diminution systématique de la méthylation de l'ADN. (ii) Afin d'évaluer la transmission des marques épigénétiques lors de la reproduction sexuée et asexuée, nous avons

comparé les épigénomes d'arbres adultes, de greffons et de semis obtenus par autofécondation sur une lignée modèle double-haploïde. Les analyses à l'échelle de génomes complets ont montré une absence de modification globale au niveau de la méthylation de l'ADN entre un arbre adulte, un greffon et un semis de pommier. Cependant nous avons détecté des régions spécifiques présentant une différence de méthylation de l'ADN (DMRs). Fait intéressant, nous avons détecté plus de DMRs entre le semis et le greffon ou l'arbre qu'entre l'arbre et le greffon, ce qui suggère que l'arbre et le greffon sont plus proches au niveau épigénétique. Nos résultats de phénotypage, d'analyses transcriptomique et de méthylome confirment la différence entre arbre adulte et semis. Mais également que les greffons sont à l'interface entre juvénilité et maturité.

Title : Epigenetics marks transmission during sexual and asexual propagation in apple (*Malus domestica*)

Keywords : *Malus domestica*, perennial plant, DNA methylation , Sexual multiplication, asexual propagation; CRISPR-Cas9, *In vitro* transformation

Abstract : Thanks to its clonal propagation (grafting), apple tree is an interesting model for epigenetic studies. Indeed, phenotypic variability can be observed within the same variety, suggesting a potential contribution of epigenetic mechanisms. At the beginning of this thesis, the mechanisms of transmission of epigenetic marks, such as DNA methylation, were still little studied in perennial plants. To better understand how epigenetic marks are transmitted and contribute to phenotypic traits, we have defined two main axes of research: (i) Using CRISPR-Cas9 technology we have created apple mutant lines in order to obtain a systematic decrease in DNA methylation. (ii) In order to assess the transmission of epigenetic marks during sexual and asexual reproduction, we compared the epigenomes of adult trees, grafts and seedlings obtained

by self-fertilization on a model doubled-haploid line. Analyses at the whole epigenome level indicated that there is no overall change in DNA methylation between an adult tree, a grafted tree and an apple seedling. However, we were able to detect specific regions that are differentially methylated (DMRs). Interestingly, more DMRs were detected between seedlings and grafted tree or adult tree than between the tree and grafted tree, suggesting that the tree and grafted tree are closer epigenetically. Our phenotypic, transcriptomic and methylome analysis results confirm the difference between adult trees and seedlings. They also confirm that grafted tree is at the interface between juvenility and maturity.