# Génétique et Epigénétique du phénomène de paramutation chez le maïs

Juliette Aubert

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

**En génétique et amélioration des plantes**

**École doctorale GAIA (N° 584)**

**Unité de recherche DIADE (UMR 232) Diversité, Adaptation et Développement des Plantes**

# Génétique et épigénétique du phénomène de paramutation chez le maïs


**Présentée par Juliette AUBERT**
**Le 17 décembre 2021**

**Sous la direction de Daniel GRIMANELLI**


**Devant le jury composé de**

| | |
|---|---|
| Moussa BENHAMED, Prof, DGG, IPS2, Université de Paris-Saclay | Rapporteur |
| Fréderic PONTVIANNE, CR, LGDP, Université de Perpignan | Rapporteur |
| Maike STAM, Associate Prof, Swammerdam Institute for Life Sciences, Université d'Amsterdam | Examinateur |
| James TREGEAR, DR, DIADE, IRD, Université de Montpellier | Président / Examinateur |
| Olivier LEBLANC, CR, DIADE, IRD, Université de Montpellier | Co-encadrant |
| Daniel GRIMANELLI, DR, DIADE, IRD, Université de Montpellier | Directeur de thèse |

UNIVERSITÉ DE MONTPELLIER

# Index

## Résumé

La paramutation est un changement d'expression méiotiquement et mitotiquement stable, résultant de l'interaction entre certains allèles. Ce phénomène a été observé à quatre loci chez le maïs, mais les mécanismes cellulaires et moléculaires qui sous-tendent le phénomène restent largement inconnus. Les acteurs de la paramutation précédemment décrits codent pour des composants de la voie de RNA-directed DNA methylation (RdDM) qui participent à la biogenèse des petits ARN interférents de 24 nucléotides (24-nt siRNA) et des longs ARN non codants. En combinant de l'immunolocalisation, de l'immunoprécipitation et le séquençage de petits ARNs, nous décrivons ARGONAUTE104 (AGO104) comme membre du complexe effecteur RdDM chargé de guider les 24-nt siRNAs vers leur cible ADN pour créer une méthylation de novo. Nous apportons ensuite la preuve que AGO104 est impliqué dans la paramutation au locus *b1* chez le maïs en utilisant le séquençage de petits ARN et l'introgression d'un mutant *ago104* dans une population paramutagénique. Par la suite, nous avons souhaité déterminer si la paramutation est un mécanisme de silencing global chez le maïs en croisant des lignées génétiquement distantes (B73, M37W et M162W) et en séquençant leur ARN messager foliaire. Nous avons identifié 147 gènes qui répondent à tous les critères de la paramutation car leur niveau d'expression est identique à celui du parent faiblement exprimé, même en l'absence de l'allèle parental d'origine. Ces gènes couvrent diverses fonctions chez le maïs, mais ne présentent pas de différences significatives de densité d'éléments transposables ou de densité de petits ARN. Avec ces 147 nouveaux candidats à la paramutation, nous soutenons que ce phénomène est plus commun que précédemment décrit chez le maïs. Enfin, des travaux antérieurs utilisant la technologie 3C (Chromosome Conformation Capture) ont montré que sept répétitions en tandem situées en amont du gène b1 sont impliquées dans la formation de boucles qui régulent potentiellement la paramutation. Pour mieux comprendre l'implication des interactions chromatiniennes dans la paramutation, nous avons utilisé la capture de la conformation des chromosomes circulaires (4C) sur des tissus somatiques de plantes qui sont, et ne sont pas, capables de paramutation (respectivement *B'* et *B'* mutant pour *mop1-1*). Nous avons cherché des interactions en *trans* qui sont spécifiques à l'allèle paramutagénique, afin d'identifier des interactions chromatiniennes liées à la paramutation.

**Mots clés** : AGO104, *booster1* (*b1*), Circular Chromosome Conformation Capture (4C), Epigénétique, Génétique, Paramutation, RNA-directed DNA Methylation (RdDM), RNAseq, siRNA, *Zea mays*

## Résumé court

La paramutation est un changement d'expression méiotiquement et mitotiquement stable, résultant d'une interaction entre allèles spécifiques. Nous décrivons ARGONAUTE104 (AGO104), un effecteur du RdDM qui est en charge de guider les 24-nt siRNAs vers leur cible pour la méthyler de novo. Nous apportons ensuite la preuve que AGO104 est impliqué dans la paramutation au locus b1 chez le maïs. Nous avons également créé une population de backcross à partir de lignées distantes et identifié 147 gènes qui répondent à tous les critères de la paramutation mais qui ne semblent pas être régulés par les acteurs connus de la paramutation. Nous avons enfin utilisé la technologie 4C pour déterminer le rôle des contacts chromatiniens dans la paramutation. Nous avons cherché des contacts chromatiniens entre régions lointaines qui sont spécifiques à un allèle paramutagénique afin d'identifier leur rôle dans la paramutation. Ce travail a permis de répondre à plusieurs questions cruciales pour la compréhension des causes et des mécanismes de la paramutation.

## Abstract

Paramutation is defined as the meiotically and mitotically heritable change in expression resulting from the interaction between specific alleles. This phenomenon has been observed at four loci in maize, but the cellular and molecular mechanisms underlying the paramutation phenomenon remain largely unknown. Previously described actors of paramutation encode components of the RNA-directed DNA-methylation (RdDM) pathway that participate in the biogenesis of 24-nt siRNAs and long non-coding RNAs. We uncover ARGONAUTE104 (AGO104) as a member of the RdDM effector complex that is in charge of guiding 24-nt siRNAs to their DNA target to create *de novo* DNA methylation, by combining immunolocalization, immunoprecipitation and small RNA sequencing. We then provide evidence that AGO104 is involved in paramutation at the *b1* locus in maize using small RNA sequencing and reverse genetics. We then tried to unravel the impact of paramutation on evolution by determining the extent of the phenomenon on maize. We created a backcross population from distant inbred lines (B73, M37W and M162W) and sequenced their leaf messenger RNA. We identified 147 genes that meet all criteria of paramutation as their silencing is stable through meiosis, and the newly silenced genes in turn silence active genes through successive backcrosses. These genes cover diverse functions in maize, but carry no significant differences in TEs or small-RNA density, and do not appear to be regulated by previously described actors of paramuation. With these 147 new candidates to paramutation, we argue that this phenomenon is more common than initially expected in maize. Previous work using 3C technology (Chromosome Conformation Capture) showed that seven tandem repeats located upstream of the paramutable *b1* gene are involved in the formation of loops that are potential regulators of paramutation. To determine the involvement of chromatin contacts in paramutation, we used Circular Chromosome Conformation Capture (4C) on somatic tissues from plants that are, and are not, capable of paramutation (respectively *B'* and *B-I*). We searched for *long-range* contacts that are allele-specific to identify paramutation-related chromatin interactions.

**Key words** : AGO104, *booster1* (*b1*), Circular Chromosome Conformation Capture (4C), Epigenetics, Genetics, Paramutation, RNA-directed DNA Methylation (RdDM), RNAseq, siRNA, *Zea mays*

## Short Summary

Paramutation is defined as the meiotically and mitotically heritable change in expression resulting from the interaction between specific alleles and its mechanisms are poorly understood. We uncover ARGONAUTE104 (AGO104) as a member of the RdDM effector complex that is in charge of guiding 24-nt siRNAs to their DNA target to create *de novo* DNA methylation. We then provide evidence that AGO104 is involved in paramutation at the *b1* locus in maize. We also created a backcross population from distant inbred lines and identified 147 genes that meet all criteria of paramutation but do not appear to be regulated by known actors of paramutation. We finally used Circular Chromosome Conformation Capture (4C) to determine the involvement of chromatin contacts in paramutation. We searched for *long-range* contacts that are allele-specific to identify paramutation-related chromatin interactions. This work answered several crucial questions towards understanding the causes and mechanisms of paramutation.

## Remerciements

La passion pour les plantes n’est probablement pas génétique, mais dans ma famille ça coule dans le sang. Je remercie mes deux grands-pères de m'avoir transmis toute jeune leur passion pour les plantes. Papi Pierre tenait un magasin de graines d'élite Clause rue Lafayette à Paris. Il vendait des plantes, des graines, des produits de traitements, de la terre. Et il se passionnait de botanique, quitte à en faire des soirées diapositives interminables de ses innombrables « photos de voyages » (appelons les « photos de fleurs » plutôt…).

Papy Jean, lui, est issu d’une famille d’agriculteurs, mais il n’est pas de la vieille école, ça non ! Il a fait des études en chimie agricole, et il a compris rapidement que les semences hybrides (de maïs, quoi d’autre ?) allaient permettre de nourrir les français. Quelle ne fût pas sa surprise, quand il a découvert ma passion pour les plantes, pour l’agronomie, et pour finalement faire une thèse en génétique du maïs ! Ah la vigueur hybride, j’en ai entendu parler bien avant de commencer mes études…

Mes respects à ma grand-mère Maguy qui a débuté sa thèse il y a 6 ans maintenant, à Barbara McClintock qui m’a donné l’envie d’étudier la génétique, et à Emmanuelle Charpentier, qui m’a inspirée pour pousser plus loin ma carrière dans la recherche, et sans qui je n’aurais pas fait de thèse. Quelques illustres modèles féminins, c’est bien plus utile qu’il n’y paraît pour les jeunes filles et les jeunes femmes.

## Equipe et collaborateurs

Merci Daniel Grimanelli pour ton encadrement et ta confiance en moi. Merci pour toute la liberté que tu m'as laissé prendre sur le sujet de thèse, dès le début. Tu m'as tout de suite encouragée à fonctionner en autonomie, tout en me guidant dans la bonne direction. Tu t'es aussi TOUJOURS rendu disponible pour mes questions. Tu m'as encouragée à aller voir le plus de séminaires possibles et à lire le plus d'articles possibles, dans mon domaine mais aussi dans tout autre domaine scientifique qui m'intéresse. C'est à mon avis l'encadrement dont j'avais besoin pour me lancer et m'épanouir dans ce métier.

Olivier Leblanc, merci à toi pour ton sauvetage à un moment de grande détresse. Tu as pris de ton temps pour me remotiver quand j'en ai eu besoin, malgré des circonstances très difficiles pour nous tous. Et tu m'as ensuite poussé hors de ma zone de confort pour terminer mon projet dans les meilleures conditions possibles. Tu t'es rendu très disponible, que ce soit pour mes questions ou pour les interminables relectures et corrections. J'admire énormément ton sens de la diplomatie, qui est un art que je ne maîtrise pas, brute de décoffrage comme je suis.

Thank you Stefan Grob for taking such a big part in my project without even being notified. We started chatting for your expertise in 4C technology and in a blink of an eye I spent almost 2 months in your lab, under your supervision, and I wrote an entire chapter about the work we did together. You are a great teacher and a super enthusiast researcher, always willing to try new experiments to test hypothesis. It was great to talk to you when I needed new perspectives on a subject.

Merci à toi Caroline Michaud pour tes oreilles attentives, tes conseils qui me sont toujours utiles, et bien sûr merci de m'avoir aidé sur la dernière ligne droite. Merci à Daphné Autran et Mathieu Ingouff de vous être rendus disponibles pour des conseils, des corrections et des relectures quand j'en ai eu besoin. Merci aussi à Julien Serret et Cédric Mariac pour les mains ingénieuses qui m'ont sorties de mon pétrin de fin de thèse.

## Famille

Merci à Matthieu, mon cher et tendre, pour ses cookies, ses tartes au citron, ses pancakes et son soutien infini et sans faille (même s'il m'a forcée à regarder le seigneur des anneaux ET le hobbit).

Merci à Maman, à Sébastien, à Clémentine, à Cécile et à Mayur de m'avoir écouté me plaindre très souvent. Merci à Élisa pour ses coups de pieds au cul et ses speechs motivationnels (quoi d'autre ?). Merci à Papa qui m'a proposé de « corriger mon rapport » jusqu'au dernier jour.

**Les copains**

Coucou les copains

Coucou Elvira, coucou Inès et coucou Daniel(ito). Vous avez tous les trois participé de manière très différente à mes trois années de thèse, mais vous partagez ici le même remerciement : merci.

Coucou Charlotte (work buddy, Les Castors buddy, party buddy, chasse aux champignons et randonnée buddy)

Coucou Marlène (vegan-not-vegan buddy, also lunch buddy, party buddy and Jägermeister buddy)

Coucou Léo Paul (lunch buddy and DJ buddy)

Coucou Carlos (swimming buddy)

Coucou Adrian (Tarzan buddy, beer buddy)

Coucou Héléna (river buddy, beer buddy, and also Jägermeister buddy)

Coucou Pablo (Dr Affortit buddy, peintre de bureau buddy (**Fig. 1**))


**Figure 1**. Œuvre d'art par Pablo Affortit, doctorant à l'IRD, DIADE, dans l'équipe CERES.

**Membres du jury et du comité de suivi**

I am grateful that Moussa Benhamed and Frédéric Pontvianne accepted to evaluate my manuscript and to be present at my PhD defense. Many thanks also to Maike Stam and James Tregear for accepting to evaluate my PhD defense.

Merci à mon référent Antoine Martin d'avoir été à l'écoute, et aux membres du comité de suivi pour leurs nombreux conseils, Christophe Tatout, Alain Ghesquière and Thierry Lagrange.

**Financements**

My PhD was funded by a state scholarship and the University of Montpellier. Most of my work was made possible by ANR CHROMOBREED (ANR-18-CE92-0041). I also received a fellowship from COST INDEPTH for a short term scientific mission (STSM) to UZH.

# Thesis introduction

I first present an overview of the main epigenetic mechanisms in maize, and their regulation during the reproductive cycle to explain transgenerational silencing phenomena. I then present more specifically paramutation in maize, as well as major unanswered questions in the field that I will try to answer in the manuscript.

## Introduction index

## What is epigenetics?

Many definitions were proposed over the years for what epigenetics is, and what notions it covers. The term was first coined in the early 1940's by Conrad Waddington, a developmental biologist interested in embryology, to describe the developmental phenomenon that directs cell specialisation during embryogenesis, because a single genome can give rise to different tissues (**Fig 1**) (Waddington, 1939; Waddington, 1940). This large concept , emphasizing the complex interplay between phenotype and genotype, was refined in 1958 as a phenomenon that enables gene expression patterns to be maintained through cell division (mitotically stable) (Nanney, 1958). Finally, the notion of chemical modifications that modify gene expression, and that are heritable through mitosis was added in 1975 (Holliday and Pugh, 1975; Riggs, 1975), and was complemented with the possibility of meiotic inheritance of epigenetic modifications over the years (Riggs et al., 1996). For clarity, the definition I selected for this PhD thesis is the most famous one, as it gathers the most important features of epigenetics: "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Riggs et al., 1996).

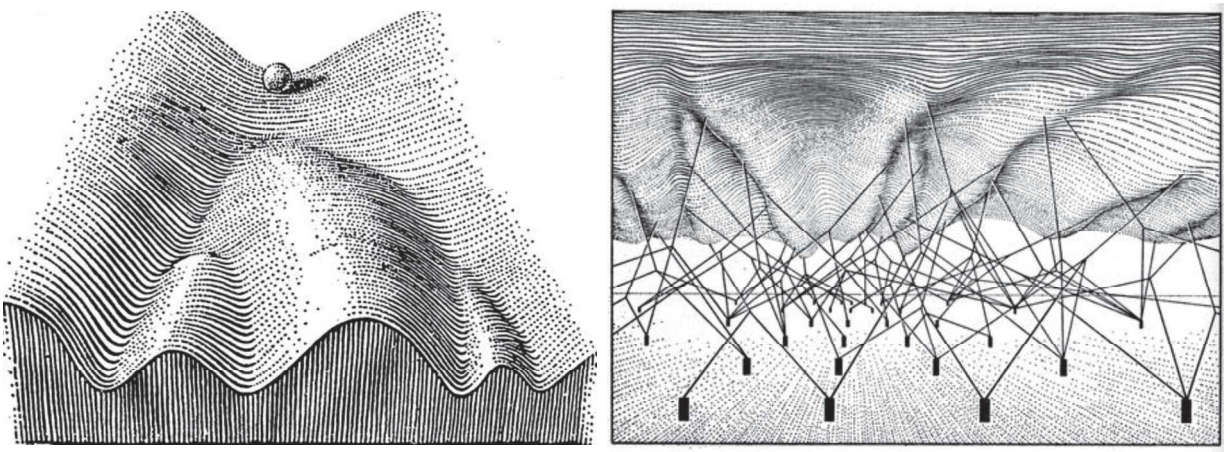


**Figure 1.** Epigenetic landscapes as described by Waddington (Waddington, 1940). A single cell can have multiple fates, depending on its direction at crucial developmental paths. Drawing on the left illustrates the different paths that a cell can take through its development, which are strongly directed by the landscape. Drawing on the right illustrates how the actions of the genome (black rectangles) can shape the landscape and direct cell fate.

## Epigenetics studies in plants and maize

Although epigenetics was described in most eukaryotes, plant genetics and epigenetics was at the center of great amount of research over the years for several considerations. First of all, plants show plethora of phenotypes, which are easily tracktable. Plants are simple to transform and genetically modify which makes both forward and reverse genetics easily conceivable in such organisms (reviewed

in (Pikaard and Mittelsten Scheid, 2014)). Plants were also often studied for their varied reproduction strategies. Most plants can be vegetatively propagated, which means that they only use mitosis in the process of reproduction. Some plants can produce seeds without meiosis (apomeiosis) nor fertilisation (parthenogenesis), a process commonly called apomixis that leads to offspring genetically identical to the mother plant (Albertini et al., 2019). Both vegetative reproduction and apomixis are crucial when studying the transmission of epigenetic features through generations. However, plants are commonly capable of polyploidy although the precise percentage is unknown, which renders any routine genetics analysis arduous. Thus, not all plants are simple models for epigenetics studies, and the choice of the studied model must consider multiple factors such as its reproductive strategy, its ploidy, and the size of its genome regarding the research question.

In plant biology and the advent of molecular biology during the 90's, the reference model has been *Arabidopsis thaliana*. It was selected as model because it is a small plant producing many seeds in a short life cycle, and which responds well to mutagenesis (nowadays, mutants are available in public collection for most of its genes) (reviewed by (Page and Grossniklaus, 2002; Pikaard and Mittelsten Scheid, 2014)). When it comes to genetics and epigenetics, *A. thaliana* is a good model because it self-pollinates (which is convenient for any genetic screening or when trying to isolate a homozygous allele), and has a small genome that was entirely sequenced. However, its small genome (135 Mb) with only 20% of transposable elements or non-coding sequences (Slotkin and Martienssen, 2007) is not representative of the enormous genetic diversity of the plant kingdom. Maize, by constrast, has a big genome (2.4 Gb), with around 80% of transposable elements (Jiao et al., 2017)(**Fig. 3**). Maize has been a central model for research in plant genetics for almost a century, since Mendel used it to validate its research in peas (reviewed in (Coe, 2001)). Transposable elements (McClintock, 1950), imprinting (Kermicle, 1969) and paramutation (Brink, 1956) are some examples of major discoveries in genetics that were made possible using the maize model. The separation of male and female reproductive organs makes it possible to control crosses, a major interest in genetics. This feature was used by many geneticists, including McClintock for demonstrating her discovery of transposable elements (**Fig. 2**)(McClintock, 1950).

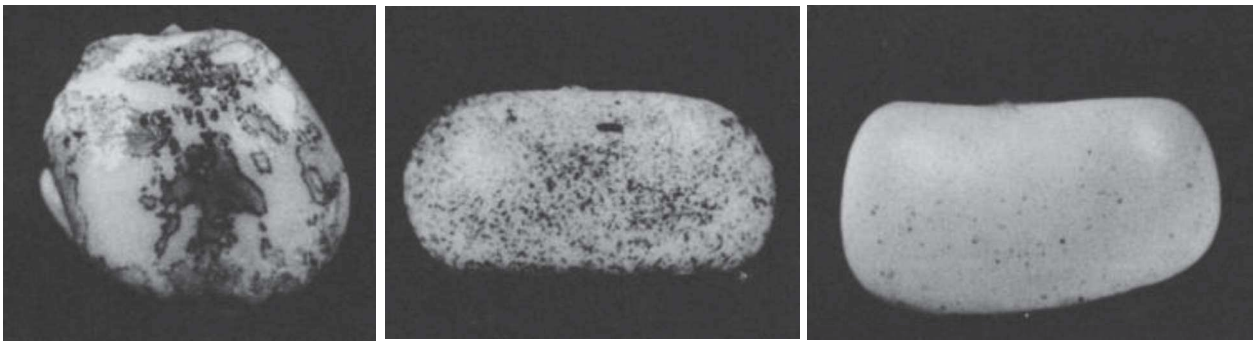


**Figure 2.** Maize kernels showing phenotype associated with TE insertion in pigmentation genes. The pictures show partial loss of pigmentation in kernels due to one, two and three TE insertions (left to right). Image adapted from McClintock (McClintock, 1951).

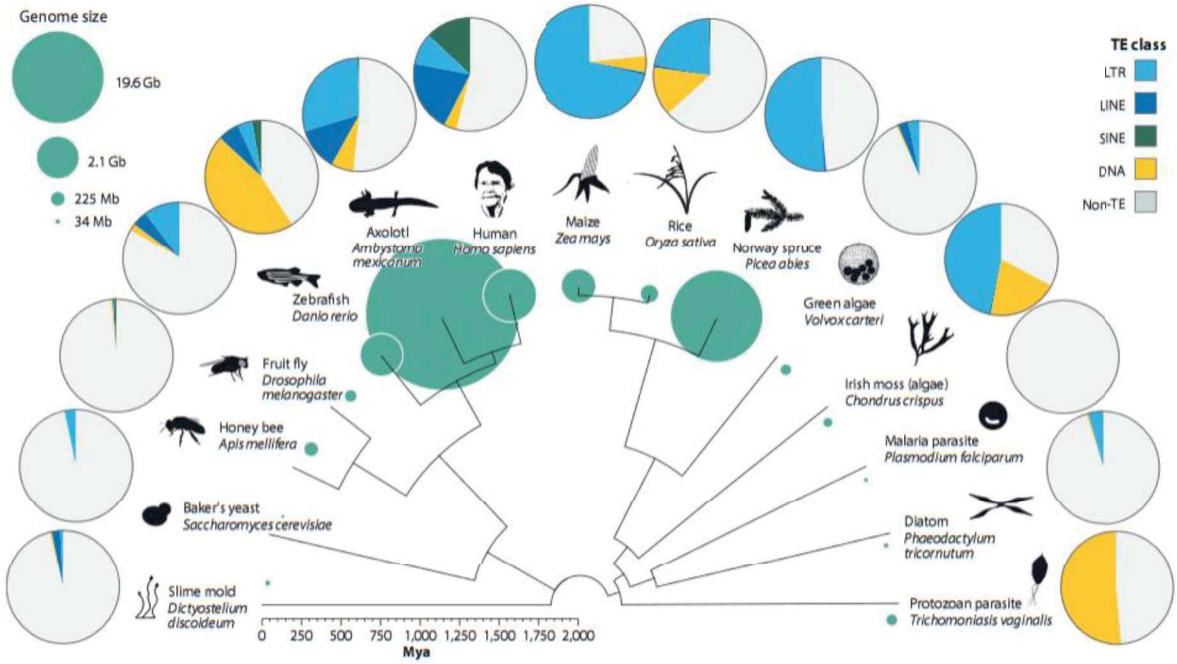



**Figure 3.** Proportion of TEs and non-TE elements in various eukaryotes genomes. Image from (Wells and Feschotte, 2020). Unfortunately, A. thaliana is not included in this figure, but has about 10% of TE in its 125 Mb genome (Kaul et al., 2000). In comparison, maize has 80% of TE in its 2.4 Gb genome elements (Jiao et al., 2017).

## Epigenetic mechanisms in maize

### DNA Methylation

DNA methylation is the addition of a methyl group on cytosine in CG, CHG and CHH contexts (with H as A, T or C). The CHG and CHH contexts of methylation only happens in plants, which adds more questions to researchers of epigenetics in plants. All three contexts of methylation are created *de novo* by the RNA-directed DNA methylation (RdDM) pathway. This pathway creates 24-nt small interfering RNA that are guided to DNA and enable to create *de novo* methylation. Once established, this

methylation can be passively lost throughout mitosis, or it can be actively maintained at symmetrical contexts (CG and CHG). CG methylation is actively maintained using *ZMMETHYLTRANSFERASE1* (*ZMET1*), a homolog of *MET1* in *A. thaliana* that acts after DNA replication and copies CG methylation patterns from the methylated strand on the newly created un-methylated strand (Li et al., 2014b). Interestingly, the increase of CG methylation within genes causes an increase of their expression, whereas increase of CG methylation within promoter sequences causes a decrease in gene expression (reviewed in (Kawashima and Berger, 2014)). CHG methylation can be not only maintained but also spread by ZMET*2* and ZMET*5*, homologs of *AtCMT3* (Li et al., 2014b). Notably, only CG and CHG methylation can be maintained through mitosis as they are symmetrical. CHH is non symmetrical and can only be added as *de novo* methylation. Interestingly, CHH methylation is mostly accumulated in close proximity to genes and enables to delimit clear transcription boundaries. Furthermore, these CHH dense regions are also associated with dense RdDM machinery, which reinforce silencing of nearby TEs (Gent et al., 2013). In *A. thaliana,* CHH and CHG methylation are also created *de novo* by CMT2, but this protein was never identified in maize (Schnable et al., 2009).

### Histone modifications

There is many different types of post-translational histone modifications, but they are always added and removed in a dynamic manner through their interaction with histone modifiers including readers, writers and erasers. Histone modifications are tissue specific and respond to environmental stresses, suggesting that they are involved in gene regulation during development and defence against stresses (Kim et al., 2008; Wang et al., 2009). Histone acetylation was found to decrease chromatin density and recruit polymerases, enabling gene expression (Hebbes et al., 1988; Wang et al., 2008; Zhang et al., 2015). On the other hand, methylation marks on histones differ between TE and genes, and their effect on gene transcription can be dramatically different depending on the methylated lysine, and on the level of methylation (Shi and Dawe, 2006; Wang et al., 2009). For instance, H3K4me3 and H3K27me3 (tri-methylation of Lysine 4 or 27 of the tail of Histone H3), are both present in euchromatin but promote respectively gene transcription and gene silencing (Shi and Dawe, 2006; Wang et al., 2009). H3K9me2 is a repressive mark, involved notably in a self-reinforcing loop of the RdDM pathway. Indeed, ZMET2 and ZMET5 are methyltransferases specialised in CHG methylation that can also recognize and bind to H3K9me2. H3K9me2 in turns recruits ZMET2 and ZMET5 which maintains surrounding CHG methylation (Du et al., 2012; Li et al., 2014b).

## Non-coding RNAs

In plants, there are many non-coding RNAs involved in epigenetics, usually categorized by their size as it is the main identifier of the mechanism they are involved with (**Fig. 4**). Small RNAs of 21 and 22-nt mainly act in Post-Transcriptional Gene Silencing (PTGS), targeting the degradation of mRNAs. In this regard, they do not interact with DNA and histone methylation epigenetic pathways (reviewed in (Cuerda-gil and Slotkin, 2016)). In *A. thaliana*, PTGS begins with the production of miRNA from specific miRNA genes encoding stem-loop structures or from other repeats loci, using POLYMERASE II (POL II) that transcribes non-silenced DNA into long non coding RNA (lncRNA). This lncRNA is processed through a combination of proteins: RNA-DEPENDANT RNA POLYMERASE6 (RDR6), DICER LIKE1 or 2 and 4 (DCL1, 2 and 4), and ARGONAUTE1 (AGO1). The produced sRNAs are guided to messenger RNA (mRNA) with sequence similarities, and their hybridization causes mRNA destruction by DCL2 and 4. As DCL2 and 4 slice RNA into 21 and 22-nt sRNA, it produces functional secondary siRNAs that will reinforce PTGS on the targeted mRNA (**Fig. 4**) (reviewed in (Cuerda-gil and Slotkin, 2016)). Another class of siRNA are 24-nt siRNA. They are produced through the RdDM (RNA directed DNA methylation) pathway, which begins with the production of the 24-nt siRNA using the RNA transcript of POL IV, and a combination of RDR2 (called MEDIATOR OF PARAMUTATION1, MOP1 in maize) and DCL3. The end of the pathway was not described in maize, but in *A. thaliana* the created 24-nt siRNAs are then guided by AGO4/6/9 to the RNA transcripts of POL V to create DNA methylation at the POL V location (**Fig. 4**) (reviewed in (Cuerda-gil and Slotkin, 2016)). In Arabidopsis, additional proteins have been identified which are involved in the targeting of POLV and the RdDM pathway, such as SHH1 (Matzke and Mosher, 2014). In addition, non-cannonical pathways have been identified, mixing PTGS and canonical RdDM pathway, for instance the specific easiRNA pathway, targeting TEs in pollen by producing siRNAs from the miRNA gene miR845, amplified by POLIV (Borges et al., 2012).

Long non coding RNA (lncRNA) are RNA longer than 200-nt that do not encode a protein, and are thought to be involved in gene regulation. These lncRNA can be produced using gene transcription machinery if they are produced nearby expressed genes, but they require other polymerases if they are intergenic lncRNA (linkRNA). Independently of this dual production mechanism, lncRNA were shown to modify gene expression and interact with chromatin remodelling complexes, suggesting an involvement in gene regulation (Li et al., 2014a).

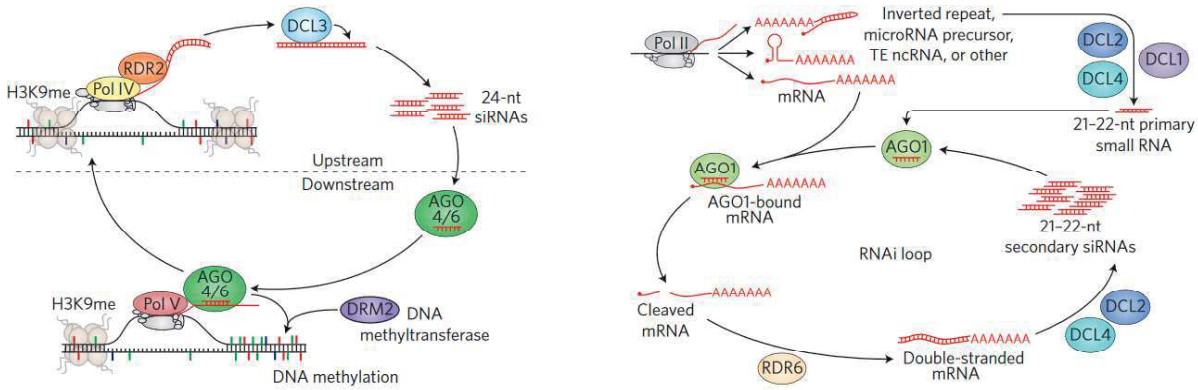



**Figure 4.** RdDM pathway (left) producing 24-nt siRNAs and PTGS (right) producing miRNAs, as described in *A. thaliana.* Image from (Cuerda-gil and Slotkin, 2016).

### Transposable Elements

TEs themselves are not epigenetic mechanisms as they are encoded in the genome, but their regulation requires various epigenetic mechanisms, which greatly influences gene regulation. TEs are defined as mobile DNA elements that are inserted in the genome and are transmitted through meiosis. Their ability to replicate (class I TE) or move (class II TE) in the genome is a potential threat to the host organism as they may insert in close proximity or within genes. Most eukaryotes have TEs but in various proportions (**Fig. 3**), and the epigenetic mechanisms used to silence TEs and prevent their mobility are highly conserved between species. Upon TE insertion, the PTGS mechanism enables to degrade TE's mRNA before the TE can be effectively silenced. Then, either the produced siRNAs hijacks the RdDM pathway to produce a strong TE silencing using methylation (reviewed in (Cuerda-gil and Slotkin, 2016)), either a homolog TE is already producing 24-nt siRNAs through RdDM and enables a rapid and efficient TE silencing through DNA methylation (reviewed in (Fultz et al., 2015)). Methylation marks are then maintained as described above, using RdDM and methyltransferases. It is worth noting that a TE inserted nearby a gene will not necessarily prevent the expression of the gene, but the strong TE silencing mechanism often spreads to surrounding regions and prevents nearby gene expression. It is then a competition between demethylases and RdDM to keep genes activated but keep TEs silenced (Gent et al., 2013; Fultz et al., 2015).

### 3D genome organisation

With advances in research technologies, it has become possible to study the 3D conformation of genomes. Topologically Associating Domains (TAD) were identified as stable genome contacts that mediate contacts between enhancers and genes, hence regulating gene expression (Sexton and Cavalli, 2015). Interestingly, it was shown that genes sharing the same TAD are more likely to have a correlated level of expression and similarly, plants with big genomes seem to organise their chromatin in

“transcription factories” that put in close proximity genes with similar transcription levels (Concia et al., 2020). Furthermore, Hi-C analysis in various tissues of maize, rice and millet showed that global euchromatin and heterochromatin domains are conserved across species and between tissues, although some tissue-specific changes were observed and associated with a change in gene regulation (Dong et al., 2020). Overall, this suggests an important role of 3D genome organisation in gene regulation. Interestingly, the span of chromatin structures may alter whether they cause gene expression or are a consequence of gene expression (reviewed in (Baroux, 2021)). It is therefore not always possible to determine the cause from the consequence even when a clear connection between chromatin conformation and gene expression is demonstrated.

## Transgenerational inheritance: resetting or transmission of epigenetic information in plants

### Plant life cycle

Four major steps take place during plant sexual reproduction: sporogenesis, gametogenesis, fertilisation and embryogenesis. In sporogenesis, the megaspore mother cell (MMC) emerges in the female reproductive tissues and the microspore mother cell (MiMC) in the male reproductive tissues, and both rapidly perform meiosis to produce the megaspore and microspore, respectively. It is worth noting that the emergence of MMC happens in adult tissues in plants, contrary to female animals that produce female gametes during foetal formation (Heard and Martienssen, 2014). Gametogenesis starts with several nuclear division of the megaspore, followed by cytokinesis, which produces the female gametophyte, composed of the egg cell (n), the central cell (2n), and antipodals and synergid cells (n). The male gametophyte is a vegetative cell (n) with one nucleus and two sperm cells (n), produced from nuclear divisions of the microspore. The vegetative cell elongates to form the pollen tube and deliver the sperm cells to the female gametophyte. The two sperm cells enable a double fertilization within the female gametophyte, of the egg cell and the central cell, which in turn produce the zygote (2n) and the endosperm (3n), respectively. Embryogenesis is the formation of a mature seed. The firstly zygote elongates, and then divides to form a heart-shaped embryo (**Fig. 5**; reviewed in (Kawashima and Berger, 2014)).

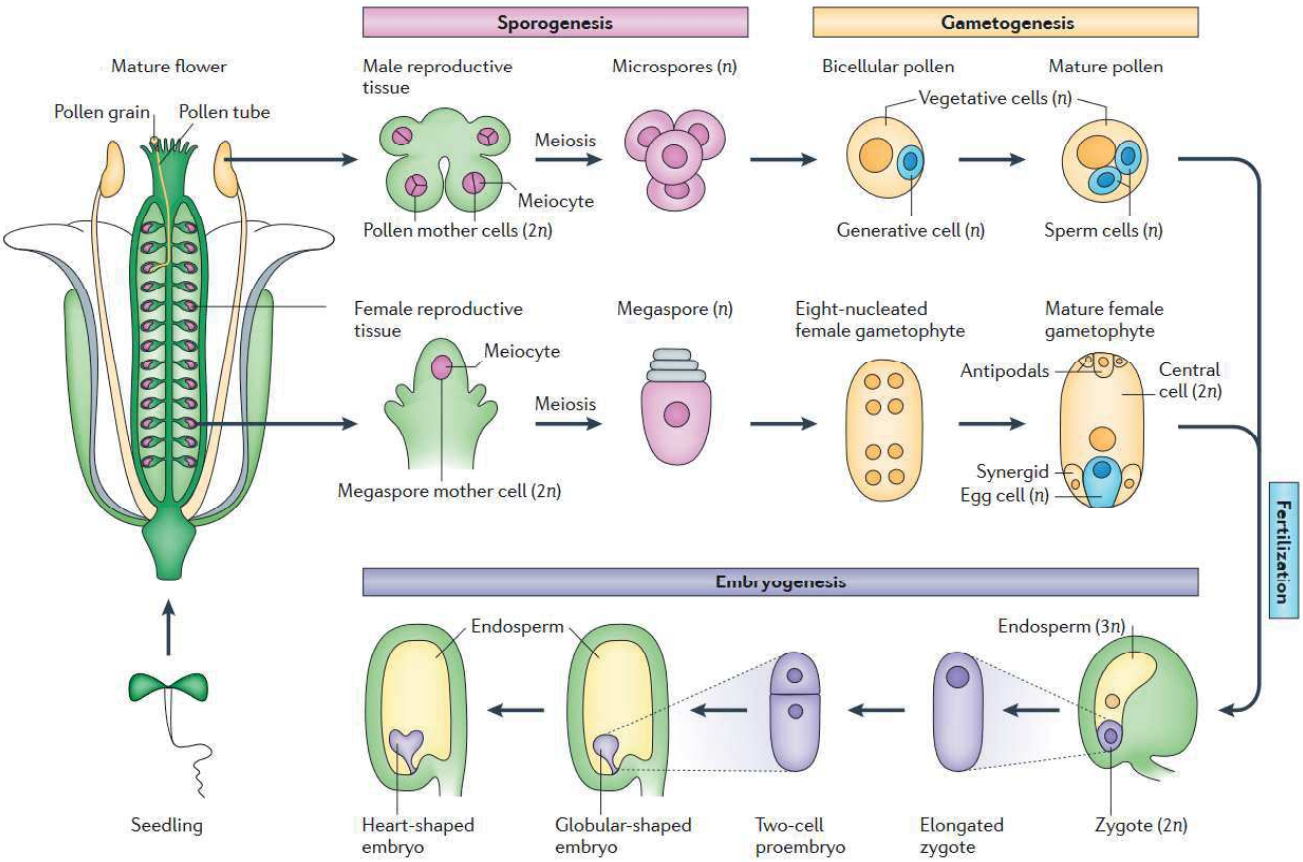



**Figure 5**. Steps of sexual reproduction in *A. thaliana*, including sporogenesis, gametogenesis, fertilisation and embryogenesis. Image from (Kawashima and Berger, 2014).

### Transgenerational inheritance vs intergenerational effects

It is first crucial to differentiate intergenerational effects from transgenerational inheritance. The intergenerational effect, or parental effect, can be caused by any stimulus (stress, poison…) and directly affects the exposed organism, as well as its current germ line. This means that female animals can perpetuate intergenerational effects in up to 3 generations, if they are pregnant during exposure to the stimuli, whereas male animals can only perpetuate intergenerational effects in up to 2 generations (Perez and Lehner, 2019). In plants, these effects are only carried over two generations because gametes are not formed until plant sexual maturity. Conversely, transgenerational inheritance are conserved through more than two generations (more than three in case of female animals) (Heard and Martienssen, 2014).

### Epigenetic regulations during sporogenesis

The visualisation of methylation dynamics in the genome was recently made possible using fluorescent reporters combined to DNA methylation binding domains (Ingouff et al., 2017). This showed that the amount of CG and CHG methylation is rather stable during meiosis in both megaspore mother cells

(MMCs) (Ingouff et al., 2017) and microspore mother cells (MiMCs) (Walker et al., 2018) in *A. thaliana*. In MMCs, CHH methylation levels decrease during meiosis and are restored afterwards (Ingouff et al., 2017), whereas CHH methylation levels are low during all MiMC stages (Walker et al., 2018). As described above, CHH methylation can only be added *de novo* because of its non symmetrical status. Such *de novo* methylation is usually performed by siRNA through the RdDM, that also performs TE silencing, which indicates that TE are activated in MMC. During MiMC sporogenesis, TE transcription is activated when CHH methylation is low, enabling the production of 24-nt siRNAs. These siRNAs bind AGO proteins to regulate gene expression and enable the formation of the male germ line (Martinez and Köhler, 2017).

### Epigenetic regulations during gametogenesis

During gametogenesis, CG and CHG contexts are highly methylated in sperm cells and silencing of TEs is lifted by a weak CHH methylation, which enables the production of epigenetically activated siRNAs (easiRNA). In contrast, the vegetative cell has a weak CG and CHG methylation and a higher CHH methylation level with activation of TE transcription in hypomethylated regions (Gehring, 2019). Although risky towards genome stability, this phenomenon likely enables the transmission of chromatin status by activating the RdDM and the PTGS pathways. This ensures a complete establishment of methylation and genome stability in the embryos (Calarco et al., 2012; Martinez and Köhler, 2017). Interestingly, CG and CHH methylation levels are stable during female gametogenesis (Ingouff et al., 2017).

### Epigenetic regulations in the embryo

During sexual reproduction, overall methylation levels vary in female and male reproductive tissues, but some parental methylation is transmitted to the embryo (Gehring, 2019). Interestingly, epigenetic reprogramming also occurs within embryos, notably at CHH contexts that are lowly methylated in the paternal gamete and heavily methylated in the mature embryo (Calarco et al., 2012; Kawashima and Berger, 2014). Although most of the parental methylation information is transmitted to embryos, post-transcriptionnally modified histones are replaced by naïve ones in the reproductive tissues, hence partially resetting chromatin conformation in the progeny (Kawashima and Berger, 2014). Such behaviour was described in *A. thaliana* where Histone H3 in the egg and sperm are replaced by naïve H3 in the zygote (Ingouff et al., 2010). In the *A. thaliana* early embryo, a global epigenetic maternal control of the paternal genome was shown to act *via* the H3K9me2 and RdDM pathways, indicating that gene expression in the embryo is still regulated by the maternal genome (Autran et al., 2011).

## An example of allele-specific epigenetic regulations in the endosperm: Genomic imprinting

In the endosperm, genomic imprinting is an example of allele-specific regulation that is not transmitted to the next generation. It is characterised by a parent-of-origin effect that determines differential gene expression of the maternal and paternal alleles in the progeny. It was first described in maize (Kermicle, 1969), and subsequently in many different organisms. In maize, imprinting was described in over 500 genes in the endosperm (Kermicle and Alleman, 1990; Waters et al., 2013). This phenomenon was linked to a weaker DNA methylation on the maternal allele, as well as histone methylation (only H3K27me3 was identified, as a silencer of gene expression), leading to a favoured expression of the maternal allele over the paternal one (Haun and Springer, 2008; Satyaki and Gehring, 2017). Paternally expressed alleles and maternally silenced alleles were also identified in Arabidopsis, with however lower prevalence as compared to maternally expressed alleles (Jahnke and Scholten, 2009; Autran et al., 2011). Interestingly, 24-nt siRNAs were found to be necessary for establishment of imprinted genes and for silencing of imprinted alleles by enabling DNA methylation (Vu et al., 2013). In most studied plants, the methylated regions causing imprinting are usually transposable elements and repeats (reviewed in (Rodrigues and Zilberman, 2015)).

## Heterosis and parental alleles interactions in hybrids

The first definition of heterosis was proposed as follows: “The decrease in vigor due to inbreeding naturally cross-fertilized species and the increase in vigor due to crossing naturally self-fertilized species are manifestations of one phenomenon. This phenomenon is heterosis.” (East and Hayes, 1912). Heterosis (or hybrid vigour) is well known from farmers as it enables to create a strong progeny with high yield from two different inbred parents. This phenomenon is more intense in maize than in other studied plants like *A. thaliana* (Meyer et al., 2004) or *Solanum lycopersicum* (Semel et al., 2006). It is commonly used in maize fields for agriculture, although the genetics behind it are quite misunderstood. In both maize and tomato, heterosis happens almost systematically on traits regulating reproduction, while it happens more rarely for other traits (Flint-Garcia et al., 2009). Genetic distance between the parents seems to increase the occurrence of heterosis in non-reproductive traits (Flint-Garcia et al., 2009). Given the known crucial role of siRNAs in trans-allelic regulation, the density of siRNA was studied in many inbred lines and their hybrid progeny. Within genes affected by heterosis, the density of expressed siRNAs did not vary when compared to either parental alleles. However, the regulatory regions of these genes in hybrids produced varying amounts of siRNAs (Crisp et al., 2020). This suggests that siRNAs act independently in each region, or that their expression is a consequence of the allelic differential gene transcription observed in heterosis, rather than its cause. Similar study was conducted in *A. thaliana* and showed that genes under heterosis regulation and their flanking

regions produced less 24-nt siNRA than their parents. Several loci also showed decreased DNA methylation levels, indicating a link between DNA methylation, 24-nt siRNA production and heterosis (Groszmann et al., 2011). These features are reminiscent to those involved in paramutation, the object of this PhD work, and another epigenetic phenomenon that involves interactions between homologous alleles and causes a change in its progeny's gene expression. It was proposed that paramutation plays a role in heterosis to regulate gene expression in the F2s and subsequent generations (Springer and McGinnis, 2015).

## Paramutation

Paramutation is a specific phenomenon that involves the trans-silencing interaction between homologous alleles responsible for heritable changes in gene expression of an highly expressed allele by a weakly expressed one. This trans-silencing is stable in subsequent generations, even in the absence of the original weakly expressed allele (**Fig. 6**). However, some paramutation loci show spontaneous reversion to their highly expressed state, but others, like the *b1* loci in maize, were never observed to revert to their expressed state (reviewed in (Chandler et al., 2000)). Interestingly, both highly and weakly expressed genes are usually genetically identical and only differ by the chromatin status of some enhancer regions, which implies that epigenetics is the main regulator of this phenomenon (Stam et al., 2002a). This highly stable phenomenon is intriguing as it was described in a reduced number of plants and animals, in a reduced number of loci (reviewed in (Gabriel and Hollick, 2015)). As methylation was shown to be transmitted through meiosis in plants, it is not surprising that such phenomenon can be stably conserved through many generations (reviewed in (Kawashima and Berger, 2014)). Paramutation is established in the developing embryos (Coe, 1966), but the dynamics of the mechanisms of establishment and maintenance of paramutation are not yet fully understood, and likely happen in the vegetative tissues (Hollick et al., 1995; Chandler et al., 2000; Haring et al., 2010).

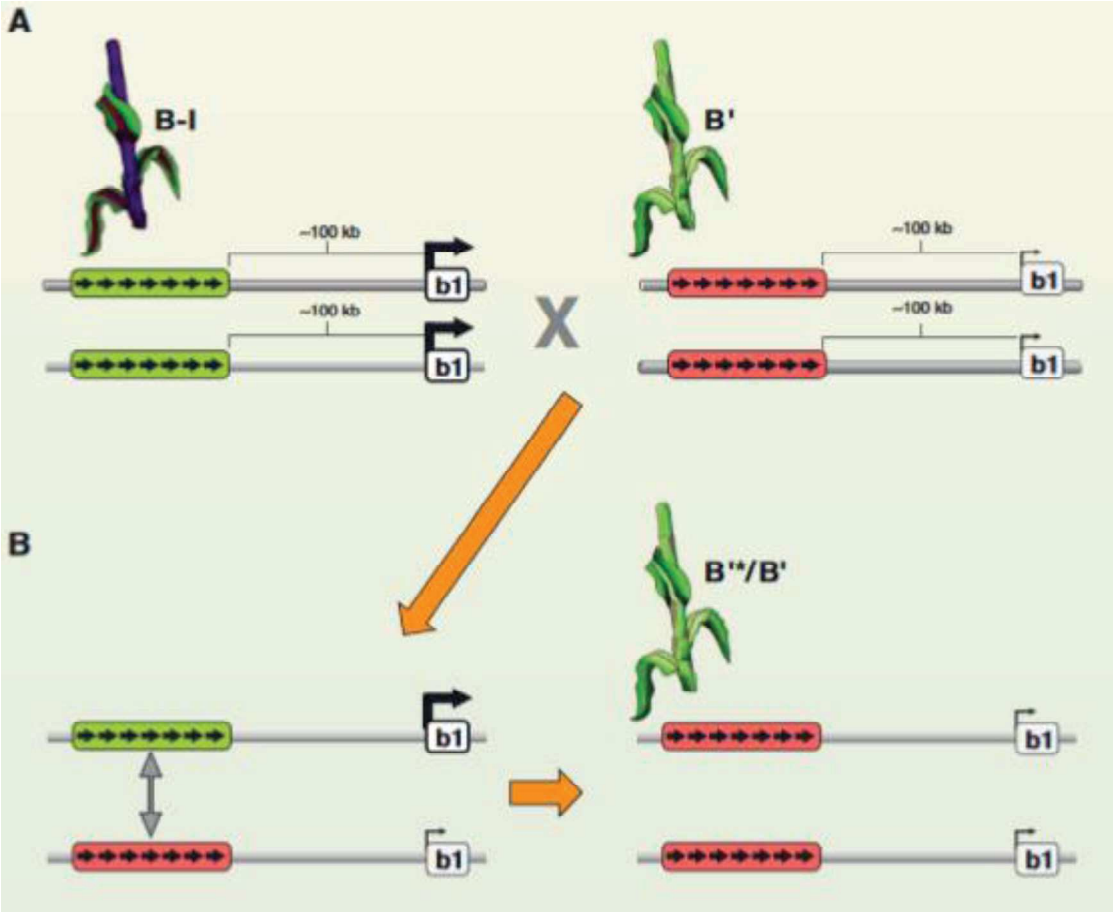



**Figure 6.** Illustration of the paramutation phenomenon at the *b1* locus in maize. The green and red rectangle represent sevent tandem repeats upstream of the *b1* locus, that are necessary to stably change the *b1* paramutation status, as well as the level of expression of *b1* and the associated stem pigmentation. Image from (Chandler, 2010).

## Description of paramutation in maize

There are few loci in maize that were described as paramutagenic. The most studied ones are *red1 (r1), booster1 (b1), pericarp color1 (p1)* and *plant color1 (pl1)*. Other loci have been described to have paramutagenic activity in maize, i.e. 145 genes show a paramutation-like transmission of their expression level in B73xMo17 RILs (Li et al., 2013).

### Paramutation at the *r1* locus

A paramutation like mechanism was described in the 1910's in pea plants (*Pisum sativum*) that showed an unsual rogue leaf phenotype with transgenerational transmission into F1 and derivative generations (**Fig. 7A**) (Bateson and Pellew, 1915). Therefore, the first discovery of a gene regulated by paramutation was done in maize at the *red colour1* locus (*r1*) in 1956 (Brink, 1956). *r1* encodes a transcription factor essential for anthocyanin biosynthesis in the aleurone layer of the seed, a readout allowing to easily

assess phenotype evolution throughout generations. Brink identified the *R-stippled* (*R-st*) paramutagenic allele with a stippled pigmentation on seeds and the *R-r* paramutable allele with a full seed pigmentation. *R-st* silences *R-r* (the paramutated allele is then called *R-r'*) and gives it a stippled pigmentation on seeds (**Fig. 7B**). This newly silenced *R-r'* allele remains silenced for a few generations, but it gradually reverts to its expressed state (*R-r*) if not exposed to the *R-st* allele again (Brink, 1956).

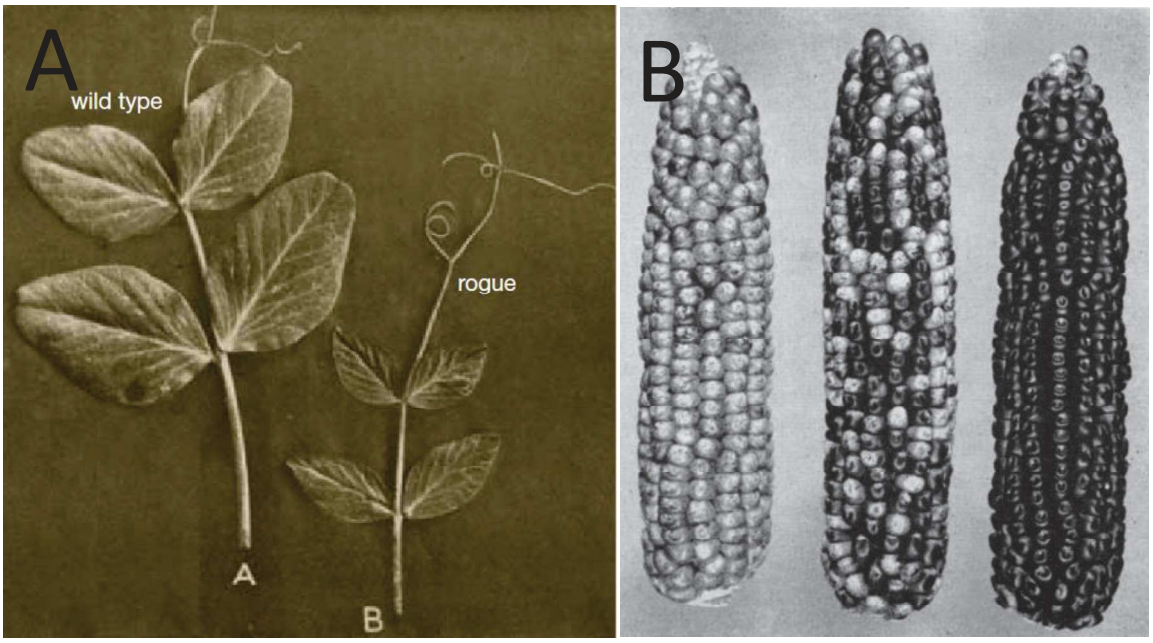



**Figure 7**. Pea and maize paramutation phenotypes. A) WT and Rogue leaf phenotype in Pea. Image taken from (Bateson and Pellew, 1915) B) Seed pigmentation in maize as a result of paramutation at the *r1* locus. From left to right, *R-st*/*R-st* kernels showing the stippled phenotype, *R-st*/*R-r'* kernels directly derived from a cross between *R-st* and *R-r* plants, and *R-r* fully pigmented kernels. Image from (Brink, 1956).

### Paramutation at the *b1* locus

The discovery of paramutation at the *r1* locus was corroborated 3 years later as the phenomenon was also identified at the *booster1* locus (*b1*). *b1* encodes a basic helix-loop-helix transcription factor activating the anthocyanin biosynthesis pathway and causes purple pigmentation which intensity depends on *b1* expression levels. Paramutation involves the *BOOSTER' (B')* allele with lightly pigmented stem and husk tissues and the BOOSTER-INTENSE (*B-I)* allele characterized by dark purple stem and husk pigmentation (**Fig. 8**) (Coe, 1959). The *b1* locus shows the same characteristics as *r1*, because the silenced, paramutagenic *B'* allele *trans*-silences *B-I* (the paramutable allele). Contrary to the *r1* locus, newly silenced *B'* alleles never revert to the *B-I* expression state. Finally, the *B-I* allele shows an intriguing behavior as up to 10% of the *b1* self-reproduced progeny shows a spontaneous silencing from the *B-I* to *B'* (Stam et al., 2002a).

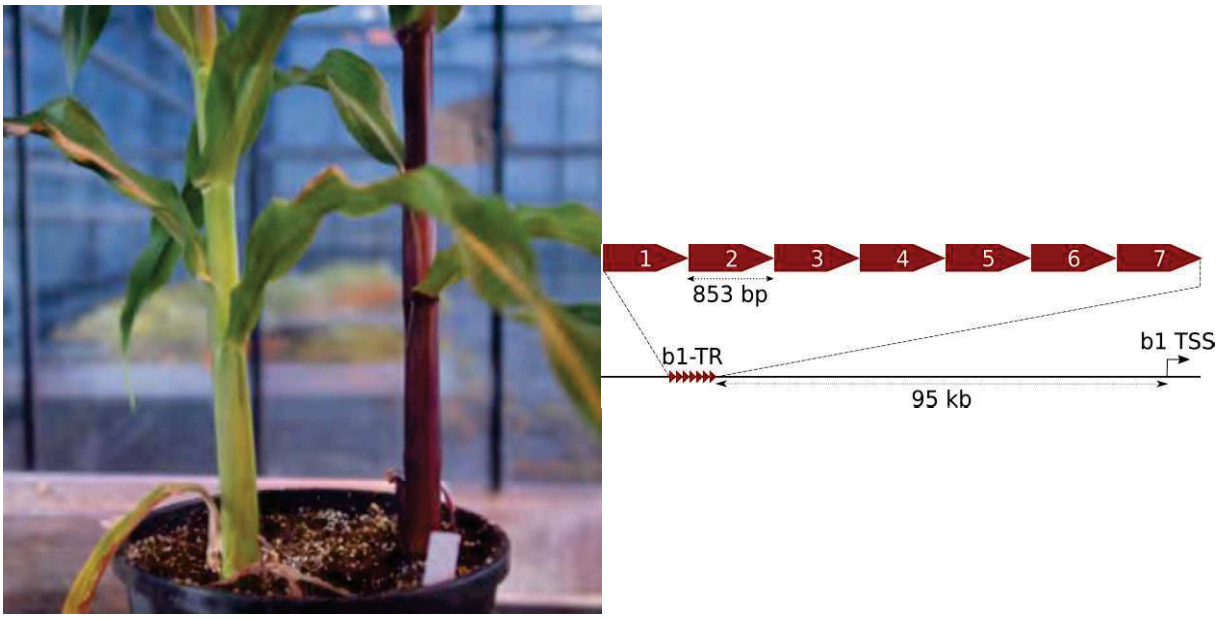



**Figure 8**. Stem pigmentation of the two alleles involved in the *b1* paramutation: *B'* on the left and *B-I* on the right. Both epialleles have seven genetically identical tandem repeats (*b1TR*) inserted 100 kb upstream of the b1 gene. These *b1TR* only differ between *B'* and *B-I* only by their chromatin state (Stam et al., 2002a).

### Which definition for the paramutation phenomenon?

A few years after the discovery of *b1* paramutation, another paramutation-like phenomenon was described at the *sulfurea* (*sulf*) locus in tomato (*Lycopersicon esculentum*) (Hagemann, 1969), that involved leaf chlorose transmitted with a paramutation-like behaviour. Depending on the writer, this phenomenon could be called somatic conversion, conversion type phenomenon, or paramutation. Therefore, Brink, Coe and Hagemann agreed on a common name and description of the characteristics from their respective discoveries (reviewed by (Pilu, 2015)). Three steps were described as corresponding to the inheritance pattern of the paramutation phenomenon, and enabled to define paramutation. The first step is called the establishment of paramutation, and is characterised by the establishment of the stable trans-silencing of one allele on an other. The second step, maintenance of paramutation, is the conservation of the weak expression of the allele over several generations. Finally, the last step is called secondary paramutation, as silenced alleles can in turn stably silence their highly expressed homologs, regardless of the presence of the original silencing allele (reviewed in (Chandler et al., 2000)). Many other characteristics could be added over the years by many scientists to what defines a paramutable gene, e.g. genetically identical genes in both alleles and identical level of methylation in both alleles, production of 24-nt siRNAs by RdDM... However, these characteristics do not apply to all examples of paramutation in eukayotes. Therefore, I chose to select the first definition

of paramutation that uses only genetic features and is the closest to what was described when the paramutable behaviour was discovered.

### Paramutation at the *pl1* locus

Using this description of paramutation, the *purple plant1 (pl1)* locus was described a few years later. (Hollick et al., 1995). It is also a transcription factor of anthocyanin biosynthesis and causes dark purple pigmentation of several vegetative tissues. As described in 1995, it involves the *Pl-Rhoades* (*Pl-Rh*) allele with a high expression of the *pl1* gene and a dark purple pigmentation of the plants. This *Pl-Rh* allele can be silenced by the *Pl-mahogany* (*Pl-mah* or *Pl'*) allele that has a weak *pl1* expression and a light purple pigmentation (**Fig. 9**). As for *r1* and *b1*, this *Pl-mah* allele is also capable of secondary paramutation as it changes *Pl-Rh* into new *Pl-mah* without the presence of the initial *Pl-mah*. Notably, this locus has a feature that is specific to *r1* and that was never observed in *b1*, as *Pl-mah* can revert to its expressed *Pl-Rh* state if it is not exposed to another *Pl-mah*.

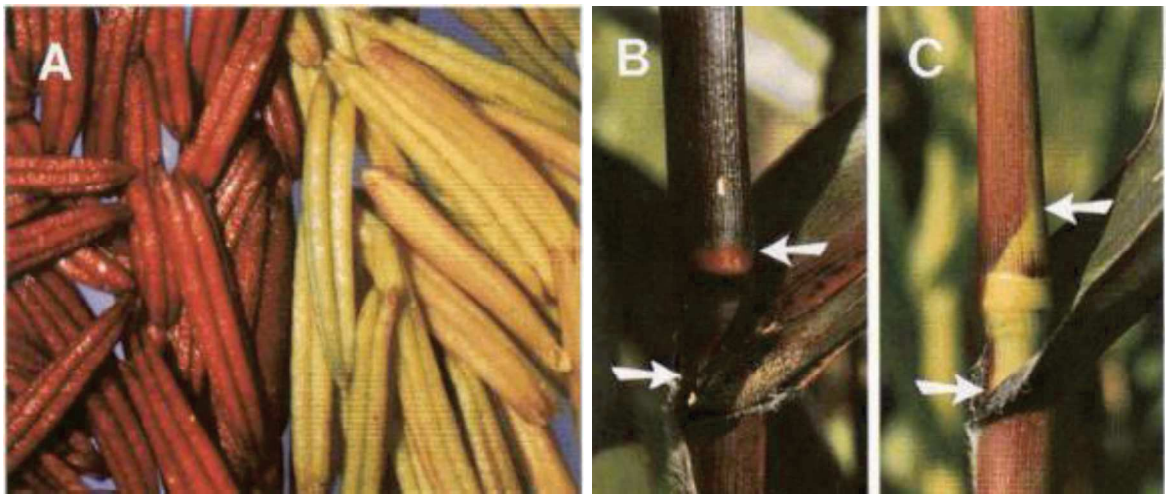



**Figure 9**. Plant pigmentation associated with *Pl-Rh* highly expressed allele (A (left) and B pictures) and with *Pl-mah* weakly expressed allele (A (right) and C pictures). Arrows indicate zones that were covered by basipetal sheath. Image adapted from (Hollick et al., 1995)

### Paramutation at the *p1* locus

Paramutation was also described at the *pericarp color1* (*p1*) locus (Das and Messing, 1994) that encodes a transcription factor controlling the synthesis of a red phlobaphene pigment which results in strong pigmentation of the cob and the pericarp. This paramutation event is recovered using transgene insertion where plants carrying a transgene of the *P1* (called *P1-rr* because of the red colour of the pericarp and cob) enhancer region produced a progeny with a decreased *P1-rr* expression, called *P1-rr'* (establishment of paramutation) (**Fig. 10**)(Sidorenko and Peterson, 2001). This *P1-rr'* state was

highly stable and maintained through generations (maintenance of paramutation), and caused the silencing of endogenous *P1-rr* alleles even when the transgene was segregated away (secondary paramutation) (Sidorenko and Peterson, 2001). Although this paramutagenic locus was produced through transgenesis, it respects all three steps that define paramutation.

**Figure 10**. *P1-rr* (left) and F1 *p1-rr'* (right) phenotypes. Adapted from (Sidorenko and Peterson, 2001).

## Mechanisms of paramutation

### Importance of enhancer regions

Many researches were conducted to understand the mechanisms involved in regulating paramutation. It was shown to be established in embryos using clonal analysis in maize zygotes (Coe, 1966), although the metastable state of *r1* and *pl1* and irreversible state of *b1* and *p1* most likely involve maintenance mechanisms occurring during the vegetative phase, as late as the 10$^{th}$ leaf (Hollick et al., 1995; Chandler et al., 2000; Haring et al., 2010). For these 4 loci, the regulation requires enhancer regions containing transposable elements (TEs) or direct tandem repeats (TR) (Kermicle et al., 1995; Sidorenko and Peterson, 2001; Stam et al., 2002b). Depending on the studied locus, these enhancers can be located upstream of the gene, distantly (up to 100 kb away for *b1*) (Stam et al., 2002b) or in close proximity to the gene (like *r1*) (Kermicle et al., 1995); or can be located both upstream and downstream of the gene (*pl1*) (Wang et al., 2017). In all cases, these regulatory elements are involved through the production of 24-nucleotide small interfering RNA (24-nt siRNA) using the RNA-directed DNA methylation (RdDM) pathway. This pathway is known to be necessary to paramutation as all mutants

of paramutation identified to date are members of RdDM (Dorweiler et al., 2000; Hollick et al., 2005; Sidorenko et al., 2009). An intensely studied example is the *b1* locus. The *b1* gene is genetically identical in both *B'* and *B-I* genetic background, and its methylation level is highly similar in both backgrounds. However, a key enhancer region of the *b1* paramutation was identified 100 kb upstream of the gene, which contains seven tandem repeats (*b1TR*) that are present in both *B'* and *B-I* backgrounds and that display different methylation profiles (Stam et al., 2002a). The paramutation status is more stable than the methylation pattern and it is likely that the methylation pattern is caused by the paramutation state (Stam et al., 2002a). Five out of the seven *b1TR* are necessary and sufficient to reproduce the *b1* paramutation, and their insertion in a transgene enables to recover most features of paramutation (Arteaga-Vazquez et al., 2010). The strong association between paramutation and regulatory sequences, especially tandem repeats, questions whether paramutation is yet another mechanism specialised to protect the genome from transgene insertion.

### Role of RdDM in paramutation

As described above, RdDM is divided in two main functions: the biosynthesis of 24-nt siRNAs, and the targeting of these siRNAs to their silencing target to produce CG, CHG or CHH methylation (H is A, T or C). The pathway starts with the transcription of regulatory elements into long non coding RNAs (lncRNAs) by POLYMERASE IV (POL IV). These lncRNAs are immediately changed into a double stranded RNA (dsRNA) by MEDIATOR OF PARAMUTATION1 (MOP1), homolog of RDR2 in *A. thaliana*. dsRNAs are targeted by DICER-LIKE3 (DCL3) and are degraded into 24-nt siRNA (Margis et al., 2006; Zhang et al., 2019). Many actors of this siRNA biosynthesis pathway were identified as necessary to paramutation in genetic screens. Among those, REQUIRED TO MAINTAIN REPRESSION6 (RMR6)/MOP3, the largest subunit of POL IV (Hollick et al., 2005; Erhard et al., 2009; Sloan et al., 2014) and RMR7/MOP2, a common subunit to POL IV and POL V (Sidorenko et al., 2009; Stonaker et al., 2009), as well as MOP1 (Dorweiler et al., 2000; Alleman et al., 2006). The homozygous *mop1-1* mutant is a striking example: all paramutated alleles (e.g. *B'*) in the *mop1-1* mutant display the phenotype of their expressed allele (e.g. *B-I* phenotype) but return to their paramutated phenotype (*B'*) when the *mop1-1* mutant is not homozygous anymore. Interestingly, these plants (*B'* plants with a *B-I* phenotype) in a *mop1-1* mutant maintained their paramutagenic activity (*B'* with a *B-I* phenotype can still silence *B-I* plants into new *B'*) (Dorweiler et al., 2000; Alleman et al., 2006). Therefore, the level of expression of the paramutagenic allele and its paramutagenicity are two different mechanisms. This was also elegantly demonstrated using *rmr6/mop3* mutants which were crossed to plants with *Pl'* allele (Hollick et al., 2005). About 36% of the progeny displayed the phenotype of the expressed allele *Pl1-Rh,* most of which returned to its paramutagenic state *pl'*, but some remained true *Pl1-Rh* in subsequent generations.

The multiplicity of mutants of the biogenesis of siRNA that disrupt paramutation shows that RdDM is required for both establishment and maintenance of paramutation. On the other hand, effectors of the second half of the RdDM pathway were only partially identified in maize, although most effectors were functionally identified in *A. thaliana*. In *A. thaliana*, the 24-nt siRNAs produced by the biogenesis part of the pathway are loaded by an ARGONAUTE (AGO) protein, AGO4, AGO6 or AGO9 (Havecker et al., 2010; Olmedo-Monfil et al., 2010) that guides the siRNAs to the lncRNA transcript of POL V (Matzke et al., 2015). This complex can bind to DNA, and recruits DOMAINS REARRANGED METHYLTRANSFERASE (DRM), DRM1 and DRM2 to induce DNA methylation at CG, CHG or CHH sites (H is A, T or C) (Law and Jacobsen, 2010; Matzke et al., 2015; Zhang et al., 2018). In maize, RMR7/MOP2 was identified as a POL V subunit, but since this subunit is common to POL IV acting upstream in the RdDM pathway, it is difficult to determine the specific role of RMR7/MOP2 and POL V in paramutation.

### *Cis* and *trans* regulation of paramutation

Paramutation is defined as a trans-silencing regulation mechanism because it is an RNA-dependent phenomenon (Stam et al., 2002b; Chandler, 2004). This was demonstrated at the b1 locus, using a transgene with the *b1TR* expressing a hairpin RNA which enabled to recover a paramutagenic allele regardless of the transgene insertion site. This demonstrates that paramutation functions in trans (Arteaga-Vazquez et al., 2010). Moreover, genome conformation and short and long range contacts are likely involved in paramutation control. Research on *b1* cis-regulation in *B'* and *B-I* epialleles proved that genome contacts are correlated with the paramutation state at *b1* (Louwers et al., 2009). Authors used Chromosome Conformation Capture (3C) experiments to reveal chromatin looping between the *b1TR*, the *b1* TSS and direct repeats located between, that are tissue-specific and allele-specific (**Fig. 11**). Although this chromatin conformation is associated with paramutation, it is unclear whether it is the cause of paramutation or its consequence. Furthermore, paramutation is known to be a trans-silencing mechanism and studying only short-range contacts might not fully cover its properties.

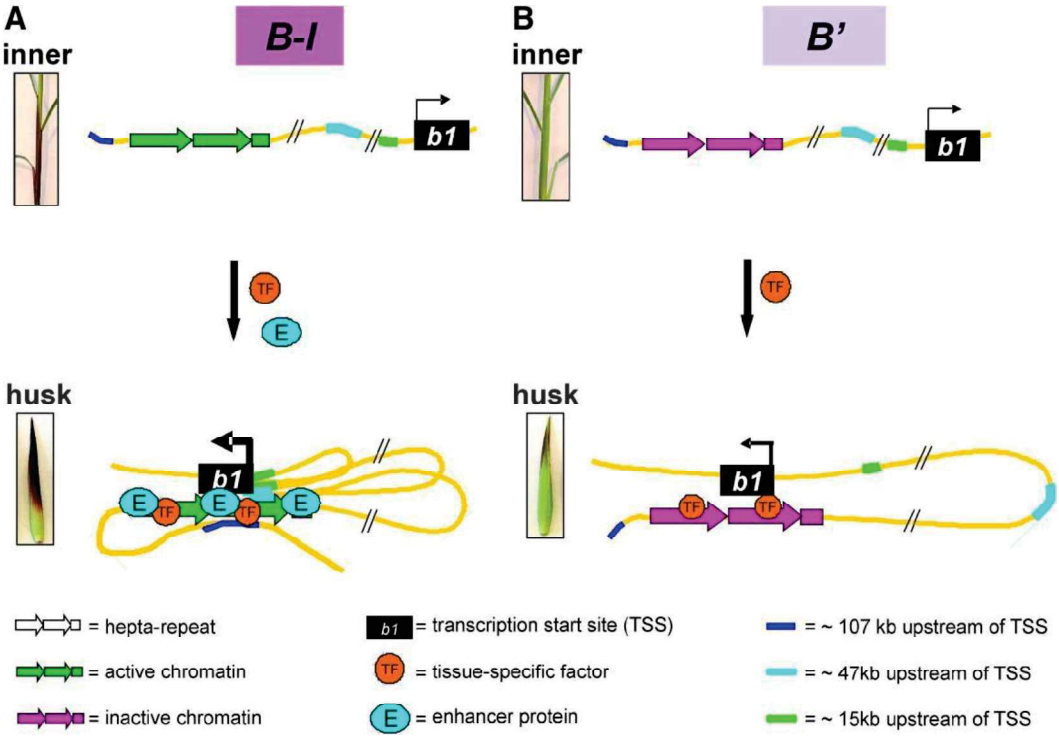



**Figure 11**. Schematic representation of short-range contacts taking place in *B-I* and *B'* epialleles. Image from (Louwers et al., 2009).

Ten KEEs were described in *A. thaliana* and are known for their high TE density and their frequent contact that all together form the KNOT structure (Grob et al., 2014). Interestingly, paramutation-like behavior was described on a transgene inserted in *A. thaliana* (Grob and Grossniklaus, 2019), correlating with a denser KNOT structure and more frequent KEE contacts. Although the mechanisms involved are unknown, this specific paramutation-like behavior can be linked to genome conformation and contacts (Grob and Grossniklaus, 2019). Whether this higher genome contact is a cause or a consequence of the phenomenon remains delicate to determine.

## Evolution of paramutation and paramutation in evolution

All four described loci of paramutation in maize involve strong tractable phenotypes with pigmentation on diverse vegetative and seed tissues. The transmission of these phenotypes in a non-mendelian manner is what raised attention to these loci in the first place. Although this behaviour was described in only four loci in maize and seems like an isolated phenomenon, similar gene behaviour was reported in many other organisms, including *Solanum lycopersicum* (Gouil and Baulcombe, 2017), *Mus musculus* (Rassoulzadegan et al., 2006), *Drosophila melanogaster* (De Vanssay et al., 2012; Ciabrelli et al., 2017), *Caenorhabditis elegans* (Sapetschnig et al., 2015), and even in humans (Bennett et al., 1997) (**Fig. 12**). Therefore, paramutation cannot be described as isolated, as it is highly conserved in eukaryotes.

However, in organisms where paramutation was described, the occurrence of the phenomenon remains limited to a small number of loci, except in *C. elegans* which uses paramutation as a global silencing mechanism to protect its genome from transgenic piwiRNA and similar sequences (Sapetschnig et al., 2015). The example of *C. elegans* is very interesting as it is the only known organism capable to both silence and activate gene expression using paramutation. It is possible that other organisms use paramutation as a global silencing mechanism, but this was never identified through previous studies focusing on isolated alleles.

**Figure 12**. Picture of paramutation at the *sulfurea* (*sulf*) locus in tomato leaf (top), and of paramutation at the *white* locus in drosophila (bottom) using the *lacZ* reporter gene to stain ovaries depending on the paramutation state. Tomato image adapted from (Gouil et al., 2016). Drosophila image adapted from (De Vanssay et al., 2012).

Epigenetics described by Waddington explains a developmental behaviour that is not described as stable through meiosis (Waddington, 1939). Epigenetics now is known to be partially stable through meiosis which raises questions about its role in evolution (Holliday, 1987). The involvement of paramutation in evolutionary processes was barely studied as this phenomenon is so far considered to be rare and not to be a global silencing mechanism. A study of the evolution of purple kernels in teosinte focused on the *r* transcription factor that was identified only as a non-functional allele in

teosinte (Hanson et al., 1996). Therefore, the silencing of the functional *r1* allele by paramutation likely first occurred in maize rather than in its ancestor. This implies that paramutation (at least at the *r1* locus) was not determinant in maize emergence but rather happened later in evolution, perhaps during inbred creation. To my knowledge, no similar study was conducted for the other loci of paramutation, and makes it difficult to infer the role of paramutation in evolution.

## Objectives and thesis structure

With this introduction, I highlighted several questions about mechanisms and occurrence of paramutation in maize. We separated these questions into three axes that are all presented in separate article-like chapters. The results for these three chapters are then discussed in the general discussion at the end of the manuscript.

The first chapter is an article untitled “AGO104 is a RdDM effector of paramutation at the maize b1 locus” that is currently under review in Frontiers in Plant Science. This paper aims at discovering the pathway that enables 24-nt siRNA to regulate paramutation. We searched for the first actor of the RdDM effector complex in maize and demonstrated its involvement in paramutation.

The second chapter is an article that will be submitted shortly to a journal for publication. It asks whether the different forms of paramutation represent an (epi)phenomenon, or a more global form of regulation of gene expression. To start answering this question, we searched for genes that undergo trans-silencing in crosses between divergent lines, with a transmission through meiosis of these silencing events. Can we identify a common feature among all identified candidates?

Finally, the third chapter is called “Discovering long-range contacts involved in paramutation using Circular Chromosome Conformation Capture”. It was also written in an article-format, although the results presented are preliminary and are not publishable as such. This chapter asks whether we can identify contacts between chromosomes during the paramutation phenomenon. Can we identify long-range contacts between alleles involved in paramutation using the 4C (Circular Chromatin Conformation Capture) technique?

My contribution for each chapter is described in the first page of each chapter.

## References


**Albertini E, Barcaccia G, Carman JG, Pupilli F** (2019) Did apomixis evolve from sex or was it the other way around? J Exp Bot **70**: 2951–2964

**Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, White J, Sikkink K, Chandler VL** (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. Nature **442**: 295–298

**Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K, Green PJ, Meyers BC, Chandler VL** (2010) RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proc Natl Acad Sci **107**: 12986–12991

**Autran D, Baroux C, Raissig MT, Lenormand T, Wittig M, Grob S, Steimer A, Barann M, Klostermeier UC, Leblanc O, et al** (2011) Maternal epigenetic pathways control parental contributions to arabidopsis early embryogenesis. Cell **145**: 707–719

**Baroux C** (2021) Three-dimensional genome organization in epigenetic regulations: cause or consequence? Curr Opin Plant Biol **61**: 102031

**Bateson W, Pellew C** (1915) On the genetics of “rogues” among culinary peas. J Genet **5**: 13–36

**Bennett ST, Wilson AJ, Esposito L, Bouzekri N, Undlien DE, Cucca F, Nisticò L, Buzzetti R, Bosi E, Pociot F, et al** (1997) Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. Nat Genet **17**: 350–352

**Borges F, Calarco JP, Martienssen RA** (2012) Reprogramming the epigenome in Arabidopsis pollen. Cold Spring Harb Symp Quant Biol **77**: 1–5

**Brink RA** (1956) A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. Genetics **41**: 872–89

**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. Cell **151**: 194–205

**Chandler VL** (2004) Poetry of b1 paramutation: cis- and trans-chromatin communication. Cold Spring Harb. Symp. Quant. Biol. pp 355–361

**Chandler VL** (2010) Paramutation ’ s Properties and Puzzles. Science (80- ) **628**: 628–630

**Chandler VL, Eggleston WB, Dorweiler JE** (2000) Paramutation in maize. Plant Mol Biol **43**: 121–145

**Ciabrelli F, Comoglio F, Fellous S, Bonev B, Ninova M, Szabo Q, Xuéreb A, Klopp C, Aravin A, Paro R, et al** (2017) Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila. Nat Genet **49**: 876–886

**Coe EH** (1966) THE PROPERTIES, ORIGIN, AND MECHANISM OF CONVERSION-TYPE INHERITANCE AT THE *B* LOCUS IN MAIZE. Genetics **53**: 1035–1063

**Coe EH** (2001) The origins of maize genetics. Nat Rev Genet **2**: 898–905

**Coe EH** (1959) A Regular and Continuing Converstion-Type Phenomenon at the B Locus in Maize. Proc Natl Acad Sci 828–832

**Concia L, Veluchamy A, Ramirez-Prado JS, Martin-Ramirez A, Huang Y, Perez M, Domenichini S, Rodriguez Granados NY, Kim S, Blein T, et al** (2020) Wheat chromatin architecture is organized in genome territories and transcription factories. Genome Biol. doi: 10.1186/s13059-020-01998-1

**Crisp PA, Hammond R, Zhou P, Vaillancourt B, Lipzen A, Daum C, Barry K, de Leon N, Buell CR, Kaeppler SM, et al** (2020) Variation and Inheritance of Small RNAs in Maize Inbreds and F1 Hybrids. Plant Physiol **182**: 318–331

**Cuerda-gil D, Slotkin RK** (2016) Non-canonical RNA-directed DNA methylation. Nat Plants. doi: 10.1038/NPLANTS.2016.163

**Das OP, Messing J** (1994) Variegated phenotype and developmental methylation changes of a maize allele originating from epimutation. Genetics **136**: 1121–1141

**Dong P, Tu X, Li H, Zhang J, Grierson D, Li P, Zhong S** (2020) Tissue-specific Hi-C analyses of rice, foxtail millet and maize suggest non-canonical function of plant chromatin domains. J Integr Plant Biol **62**: 201–217

**Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL** (2000) *mediator of paramutation1* Is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci. Plant Cell **12**: 2101–2118

**Du J, Zhong X, Bernatavichute Y V., Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A, et al** (2012) Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell **151**: 167–180

**East EM, Hayes HK** (1912) Heterozygosis in Evolution and in Plant Breeding.

**Erhard KF, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB** (2009) RNA polymerase IV functions in paramutation in Zea mays. Science (80- ) **323**: 1201–1205

**Flint-Garcia SA, Buckler ES, Tiffin P, Ersoz E, Springer NM** (2009) Heterosis Is Prevalent for Multiple Traits in Diverse Maize Germplasm. PLoS One. doi: 10.1371/journal.pone.0007433

**Fultz D, Choudury SG, Slotkin RK** (2015) Silencing of active transposable elements in plants. Curr Opin Plant Biol **27**: 67–76

**Gabriel JM, Hollick JB** (2015) Paramutation in maize and related behaviors in metazoans. Semin Cell Dev Biol **44**: 11–21

**Gehring M** (2019) Epigenetic dynamics during flowering plant reproduction: evidence for reprogramming? New Phytol nph.15856

**Gent JI, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, Dawe RK** (2013) CHH islands: De novo DNA methylation in near-gene chromatin regulation in maize. Genome Res **23**: 628–637

**Gouil Q, Baulcombe DC** (2017) Frequent paramutation-like features of natural epialleles in tomato. BMC Genomics **19**: 1–15

**Gouil Q, Novák O, Baulcombe DC** (2016) SLTAB2 is the paramutated SULFUREA locus in tomato. J Exp Bot **67**: 2655–2664

**Grob S, Grossniklaus U** (2019) Invasive DNA elements modify the nuclear architecture of their insertion site by KNOT-linked silencing in Arabidopsis thaliana. Genome Biol **20**: 120

**Grob S, Schmid MW, Grossniklaus U** (2014) Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. Mol Cell **55**: 678–693

**Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, Dennis ES** (2011) Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. Proc Natl Acad Sci U S A **108**: 2617–22

**Hagemann R** (1969) SOMATIC CONVERSION (PARAMUTATION) AT THE SULFUREA LOCUS OF LYCOPERSICON ESCULENTUM MILL. III. STUDIES WITH TRISOMICS . Can J Genet Cytol **11**: 346–358

**Hanson MA, Gaut BS, Stec AO, Fuerstenberg SI, Goodman MM, Coe EH, Doebley JF** (1996) Evolution of anthocyanin biosynthesis in maize kernels: The role of regulatory and enzymatic loci. Genetics **143**: 1395–1407

**Haring M, Bader R, Louwers M, Schwabe A, Van Driel R, Stam M** (2010) The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. Plant J **63**: 366–378

**Haun WJ, Springer NM** (2008) Maternal and paternal alleles exhibit differential histone methylation and acetylation at maize imprinted genes. Plant J **56**: 903–912

**Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC** (2010) The arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. Plant Cell **22**: 321–334

**Heard E, Martienssen RA** (2014) Transgenerational epigenetic inheritance: Myths and mechanisms. Cell **157**: 95–109

**Hebbes TR, Thorne AW, Crane-Robinson C** (1988) A direct link between core histone acetylation and transcriptionally active chromatin. EMBO J **5**: 1395–1402

**Hollick JB, Kermicle JL, Parkinson SE** (2005) Rmr6 maintains meiotic inheritance of paramutant states in Zea mays. Genetics **171**: 725–740

**Hollick JB, Patterson GI, Coe EH, Cone KC, Chandler VL** (1995) Allelic interactions heritably alter the activity of a metastable maize pl allele. Genetics **141**: 709–719

**Holliday R** (1987) The inheritance of epigenetic defects. Science (80- ) **238**: 163–170

**Holliday R, Pugh JE** (1975) DNA modification mechanisms and gene activity during development. 187–226

**Ingouff M, Rademacher S, Holec S, Šoljić L, Xin N, Readshaw A, Foo SH, Lahouze B, Sprunck S, Berger F** (2010) Zygotic resetting of the HISTONE 3 variant repertoire participates in epigenetic reprogramming in arabidopsis. Curr Biol **20**: 2137–2143

**Ingouff M, Selles B, Michaud C, Vu TM, Berger F, Schorn AJ, Autran D, Van Durme M, Nowack MK, Martienssen RA, et al** (2017) Live-cell analysis of DNA methylation during sexual reproduction in arabidopsis reveals context and sex-specific dynamics controlled by noncanonical RdDM. Genes Dev **31**: 72–83

**Jahnke S, Scholten S** (2009) Epigenetic Resetting of a Gene Imprinted in Plant Embryos. Curr Biol **19**: 1677–1681

**Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, et al** (2017) Improved maize reference genome with single-molecule technologies. Nature **546**: 524–527

**Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ, Feldblyum T, Nierman W, Benito MI, Lin X, et al** (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature **408**: 796–815

**Kawashima T, Berger F** (2014) Epigenetic reprogramming in plant sexual reproduction. Nat Rev Genet **15**: 613–624

**Kermicle JL** (1969) Androgenesis conditioned by a mutation in maize. Science (80- ) **166**: 1422–1424

**Kermicle JL, Alleman M** (1990) Gametic imprinting in maize in relation to the angiosperm life cycle. Dev Suppl 9–14

**Kermicle JL, Eggleston WB, Alleman M** (1995) Organization of paramutagenicity in R-stippled maize. Genetics 141(1) 361-372

**Kim JM, To TK, Ishida J, Morosawa T, Kawashima M, Matsui A, Toyoda T, Kimura H, Shinozaki K, Seki M** (2008) Alterations of lysine modifications on the histone H3 N-tail under drought stress conditions in Arabidopsis thaliana. Plant Cell Physiol **49**: 1580–1588

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

**Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chettoor AM, Givan SA, Cole RA, Fowler JE, et al** (2014a) Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biol **15**: R40

**Li L, Petsch K, Shimizu R, Liu S, Xu WW, Ying K, Yu J, Scanlon MJ, Schnable PS, Timmermans MCP, et al** (2013) Mendelian and Non-Mendelian Regulation of Gene Expression in Maize. PLoS Genet **9**: e1003202

**Li Q, Eichten SR, Hermanson PJ, Zaunbrecher VM, Song J, Wendt J, Rosenbaum H, Madzima TF, Sloan AE, Huang J, et al** (2014b) Genetic Perturbation of the Maize Methylome. Plant Cell **26**: 4602–4616

**Louwers M, Bader R, Haring M, van Driel R, de Laat W, Stam M** (2009) Tissue- and Expression Level-Specific Chromatin Looping at Maize b1 Epialleles. Plant Cell Online **21**: 832–842

**Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM** (2006) The evolution and diversification of Dicers in plants. FEBS Lett **580**: 2442–2450

**Martinez G, Köhler C** (2017) Role of small RNAs in epigenetic reprogramming during plant sexual reproduction. Curr Opin Plant Biol **36**: 22–28

**Matzke MA, Kanno T, Matzke AJM** (2015) RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. Annu Rev Plant Biol **66**: 243–267

**Matzke MA, Mosher RA** (2014) RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. Nat Rev Genet **15**: 394–408

**McClintock B** (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci U S A **36**: 344–355

**McClintock B** (1951) Chromosome organization and genic expression. Cold Spring Harb Symp Quant Biol **16**: 13–47

**Meyer RC, Törjék O, Becher M, Altmann T** (2004) Heterosis of biomass production in arabidopsis. Establishment during early development. Plant Physiol **134**: 1813–1823

**Nanney DL** (1958) EPIGENETIC CONTROL SYSTEMS. Proc Natl Acad Sci **44**: 712–717

**Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Arévalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP** (2010) Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature **464**: 628–632

**Page DR, Grossniklaus U** (2002) The art and design of genetic screens: Arabidopsis thaliana. Nat Rev Genet **3**: 124–136

**Perez MF, Lehner B** (2019) Intergenerational and transgenerational epigenetic inheritance in animals. Nat Cell Biol 1

**Pikaard CS, Mittelsten Scheid O** (2014) Epigenetic regulation in plants. Cold Spring Harb Perspect Biol **6**: a019315

**Pilu R** (2015) Paramutation phenomena in plants. Semin Cell Dev Biol **44**: 2–10

**Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F** (2006) RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. doi: 10.1038/nature04674

**Riggs AD** (1975) X inactivation, differentiation, and DNA methylation. Cytogenet Genome Res **14**: 9–25

**Riggs AD, Martienssen RA, Russo VE** (1996) Introduction. Epigenetic Mech gene Regul 0–4

**Rodrigues JA, Zilberman D** (2015) Evolution and function of genomic imprinting in plants. Genes Dev **29**: 2517–2531

**Sapetschnig A, Sarkies P, Lehrbach NJ, Miska EA** (2015) Tertiary siRNAs Mediate Paramutation in C.

elegans. PLoS Genet **11**: 1005078

**Satyaki PR V., Gehring M** (2017) DNA methylation and imprinting in plants: machinery and mechanisms. Crit Rev Biochem Mol Biol **52**: 163–175

**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 (80- ) **326**: 1112–1115

**Semel Y, Nissenbaum J, Menda N, Zinder M, Krieger U, Issman N, Pleban T, Lippman Z, Gur A, Zamir D** (2006) Overdominant quantitative trait loci for yield and fitness in tomato. Proc Natl Acad Sci U S A **103**: 12981–12986

**Sexton T, Cavalli G** (2015) The role of chromosome domains in shaping the functional genome. Cell **160**: 1049–1059

**Shi J, Dawe RK** (2006) Partitioning of the Maize Epigenome by the Number of Methyl Groups on Histone H3 Lysines 9 and 27. Genetics **173**: 1571–1583

**Sidorenko L, Dorweiler JE, Cigan AM, Arteaga-Vazquez M, Vyas M, Kermicle J, Jurcin D, Brzeski J, Cai Y, Chandler VL** (2009) A dominant mutation in mediator of paramutation2, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. PLoS Genet. doi: 10.1371/journal.pgen.1000725

**Sidorenko L V., Peterson T** (2001) Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize p1 gene. Plant Cell **13**: 319–335

**Sloan AE, Sidorenko L, McGinnis KM** (2014) Diverse gene-silencing mechanisms with distinct requirements for RNA polymerase subunits in Zea mays. Genetics **198**: 1031–1042

**Slotkin RK, Martienssen R** (2007) Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet **8**: 272–285

**Springer NM, McGinnis KM** (2015) Paramutation in evolution, population genetics and breeding. Semin Cell Dev Biol **44**: 33–38

**Stam M, Belele C, Dorweiler JE, Chandler VL** (2002a) Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev **16**: 1906–1918

**Stam M, Belele C, Ramakrishna W, Dorweiler JE, Bennetzen JL, Chandler VL** (2002b) The regulatory regions required for B' paramutation and expression are located far upstream of the maize b1 transcribed sequences. Genetics 162(2) 917-930

**Stonaker JL, Lim JP, Erhard KF, Hollick JB** (2009) Diversity of Pol IV function is defined by mutations at the maize rmr7 locus. PLoS Genet. doi: 10.1371/journal.pgen.1000706

**De Vanssay A, Bougé AL, Boivin A, Hermant C, Teysset L, Delmarre V, Antoniewski C, Ronsseray S** (2012) Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature **490**:

112–115

**Vu TM, Nakamura M, Calarco JP, Susaki D, Lim PQ, Kinoshita T, Higashiyama T, Martienssen RA, Berger F** (2013) RNA-directed DNA methylation regulates parental genomic imprinting at several loci in Arabidopsis. Dev **140**: 2953–2960

**Waddington CH** (1939) An Introduction to Modern Genetics. Routledge. doi: 10.4324/9781315665412

**Waddington CH** (1940) Organisers and genes. Cambridge Biological Studies. University Press, Cambridge.

**Walker J, Gao H, Zhang J, Aldridge B, Vickers M, Higgins JD, Feng X** (2018) Sexual-lineage-specific DNA methylation regulates meiosis in Arabidopsis. Nat Genet **50**: 130–137

**Wang P-H, Wittmeyer KT, Lee T, Meyers BC, Chopra S** (2017) Overlapping RdDM and non-RdDM mechanisms work together to maintain somatic repression of a paramutagenic epiallele of maize pericarp color1. PLoS One **12**: e0187157

**Wang X, Elling AA, Li X, Li N, Peng Z, He G, Sun H, Qi Y, Liu XS, Deng XW** (2009) Genome-Wide and Organ-Specific Landscapes of Epigenetic Modifications and Their Relationships to mRNA and Small RNA Transcriptomes in Maize. Plant Cell **21**: 1053–1069

**Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, et al** (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet **40**: 897–903

**Waters AJ, Bilinski P, Eichten SR, Vaughn MW, Ross-Ibarra J, Gehring M, Springer NM** (2013) Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. Proc Natl Acad Sci U S A **110**: 19639–19644

**Wells JN, Feschotte C** (2020) A Field Guide to Eukaryotic Transposable Elements. Annu Rev Genet **54**: annurev-genet-040620-022145

**Zhang H, Lang Z, Zhu JK** (2018) Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol **19**: 489–506

**Zhang W, Garcia N, Feng Y, Zhao H, Messing J** (2015) Genome-wide histone acetylation correlates with active transcription in maize. Genomics **106**: 214–220

**Zhang Z, Teotia S, Tang J, Tang G** (2019) Perspectives on microRNAs and phased small interfering RNAs in maize (Zea mays L.): Functions and big impact on agronomic traits enhancement. Plants. doi: 10.3390/plants8060170

# Chapter 1

# AGO104 is a RdDM effector of paramutation at the maize *b1* locus

The lack of precise information on effector members of the RdDM pathway in maize raised questions that I partially addressed in the first chapter: Who are the actors of the effector complex in maize and are they involved in paramutation? To answer that question, we first demonstrated that AG0104, a maize putative homolog of *A. thaliana* AGO9, is a RdDM effector by combining several approaches including Western blot analyses, immunoprecipitation followed by small RNA extraction, stem loop PCR and small RNA sequencing. We next validated AGO104 as a candidate regulator of paramutation at the *b1* locus using the previous small RNA sequencing and immunolocalization. Finally, we applied a reverse genetic approach to demonstrate that the *ago104* mutation alters paramutation at the *b1* locus. We therefore identified AGO104 as a new effector of the RdDM pathway in maize and showed its involvement in paramutation at the *b1* locus. Although other AGO proteins may also play a role in maize RdDM, these results demonstrate that the effector complex is functional in maize, as it was assumed for a long time but never proved.

**My contribution**

In this chapter, Fanny Bellegarde performed the crosses and the phenotypic analyses except pigmentation levels for which I did quantification analyses. Fanny Bellegarde also produced the immunolocalization, stem-loop PCR and Western blot experiments. However, I performed the RNA-immunoprecipitation and small RNA extraction from AGO104. I prepared the small RNA Illumina libraries and performed all subsequent bioinformatic analyses. I wrote the manuscript and presented this work at the 2019 European Maize Genetics Meeting, Montpellier, France.

This paper is currently under review in Plos One (submitted December 8, 2021).

# Chapter 1 Index

# AGO104 is a RdDM effector of paramutation at the maize *b1* locus

Juliette Aubert [1¶], Fanny Bellegarde [1#¶], Omar Oltehua-Lopez [2], Olivier Leblanc [1], Mario A. Arteaga-Vazquez [2], Robert A. Martienssen [3] & Daniel Grimanelli [1*]

[1] DIADE, University of Montpellier, CIRAD, IRD, Montpellier, France

[2] Universidad Veracruzana, INBIOTECA, Xalapa, Veracruz, Mexico

[3] Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA

[¶] These authors contributed equally to this work.

[#] Present address: Graduate School of Bioagricultural Sciences, Nagoya University, Chikusa, Nagoya, Japan

[*] Correspondence:
Daniel Grimanelli
Tel: +33 (0)4 6741 6376
Email: daniel.grimanelli@ird.fr

**Keywords:** ARGONAUTE104 (AGO104), *b1* locus, effector complex, paramutation, RNA-directed DNA methylation (RdDM), siRNA, *Zea mays*

## Abstract

Although paramutation has been well-studied at a few hallmark loci involved in anthocyanin biosynthesis in maize, the cellular and molecular mechanisms underlying the phenomenon remain largely unknown. Previously described actors of paramutation encode components of the RNA-directed DNA-methylation (RdDM) pathway that participate in the biogenesis of 24-nucleotide small interfering RNAs (24-nt siRNAs) and long non-coding RNAs. In this study, we uncover an ARGONAUTE (AGO) protein as an effector of the RdDM pathway that is in charge of guiding 24-nt siRNAs to their DNA target to create *de novo* DNA methylation. We combined immunolocalization, immunoprecipitation, small RNA sequencing and reverse genetics to, first, validate AGO104 as a member of the RdDM effector complex and, then, investigate its role in paramutation. We found that AGO104 binds 24-nt siRNAs involved in RdDM, including those required for paramutation at the *b1* locus. We also show that the *ago104-5* mutation disrupts the paramutation phenotype at the *b1* locus

and causes an intermediate stem pigmentation phenotype. Therefore, our results provide evidence that AGO104 is an additional member of the RdDM effector complex that plays a role in paramutation at the *b1* locus in maize.

## Introduction

Paramutation is defined as the meiotically and mitotically heritable change in expression resulting from the interaction between specific alleles (Brink, 1973; Arteaga-Vazquez and Chandler, 2010; Giacopelli and Hollick, 2015; Hövel et al., 2015; Hollick, 2017). This phenomenon has been observed at four loci in maize, all encoding a transcription factor mediating flavonoid biosynthesis: *red1* (*r1*), *plant color1* (*pl1), pericarp color1* (*p1*) and *booster1* (*b1*). Paramutation at *b1* is one of the best characterized systems (Stam et al., 2002a; Stam et al., 2002b; Arteaga-Vazquez et al., 2010). It involves the highly transcribed *BOOSTER-INTENSE* (*B-I*) allele causing dark pigmentation in most tissues and the *BOOSTER'* (*B'*) allele which lower expression results in light pigmentation. When *B-I* and *B'* are combined, *B'* induces the meiotically stable *trans*-silencing of *B-I* and this conversion is permanent. In addition, *B-I* converted alleles acquire *B'* paramutagenic capacity and therefore can trigger secondary paramutation events in the next generation. High transcription and full paramutagenicity (*trans*-silencing) at the *B-I* allele require the presence of at least five tandem repeats of a 853-bp sequence (*b1TR)* located ~100 kb upstream of the transcription starting site (Stam et al., 2002b; Stam et al., 2002a). The *b1TRs* produce 24-nucleotide (nt) small interfering RNAs (siRNAs) through the RNA-directed DNA Methylation (RdDM) pathway (Arteaga-Vazquez et al., 2010). Previous studies demonstrated that paramutation has an establishment phase in developing embryos, but the irreversible change from *B-I* to *B'* likely occurs during the vegetative phase, where *b1TRs* in *B-I* are gradually methylated to reach methylation levels found in *B'* (Hollick et al., 1995; Chandler et al., 2000; Haring et al., 2010). There is evidence that the RdDM pathway is critical for both establishment and maintenance of paramutation in maize (Alleman et al., 2006; Erhard et al., 2009; Sidorenko et al., 2009; Arteaga-Vazquez and Chandler, 2010; Barbour et al., 2012; Belele et al., 2013; Erhard et al., 2013).

The RdDM pathway is composed of two main functions, with the first one devoted to the biogenesis of 24-nt siRNAs and the second one, called the effector complex, employing these siRNAs as guides to establish sequence specific cytosine methylation and transcriptional repression (Fig. **1a**). In the first step, POL IV transcripts are immediately converted into double-stranded RNAs (dsRNAs) by MEDIATOR OF PARAMUTATION1 (MOP1), the homolog of RDR2 in *Arabidopsis thaliana*. DICER-LIKE3a (DCL3a) then slices these dsRNAs into 24-nt siRNAs (Margis et al., 2006; Zhang et al., 2019) which are necessary

to the effector complex to induce DNA methylation at either CG, CHG or CHH sites (where H=A, T, or C) (Fig. **1a**). Few members of the effector complex were identified in maize, although they were extensively described in *A. thaliana* (reviewed in (Law and Jacobsen, 2010; Matzke and Mosher, 2014; Matzke et al., 2015; Zhang et al., 2018)). In *A. thaliana*, it initiates with AtAGO4/6/9 (Havecker et al., 2010; Olmedo-Monfil et al., 2010), that guide siRNAs to long non-coding scaffold transcripts generated by POL V that can bind DNA (Matzke et al., 2015; Parent et al., 2021). The complex then partners with DOMAINS REARRANGED METHYLTRANSFERASE (DRM), DRM1 and DRM2, to enable DNA methylation in all sequence contexts (Law and Jacobsen, 2010; Matzke et al., 2015; Zhang et al., 2018) (Fig **1A**).

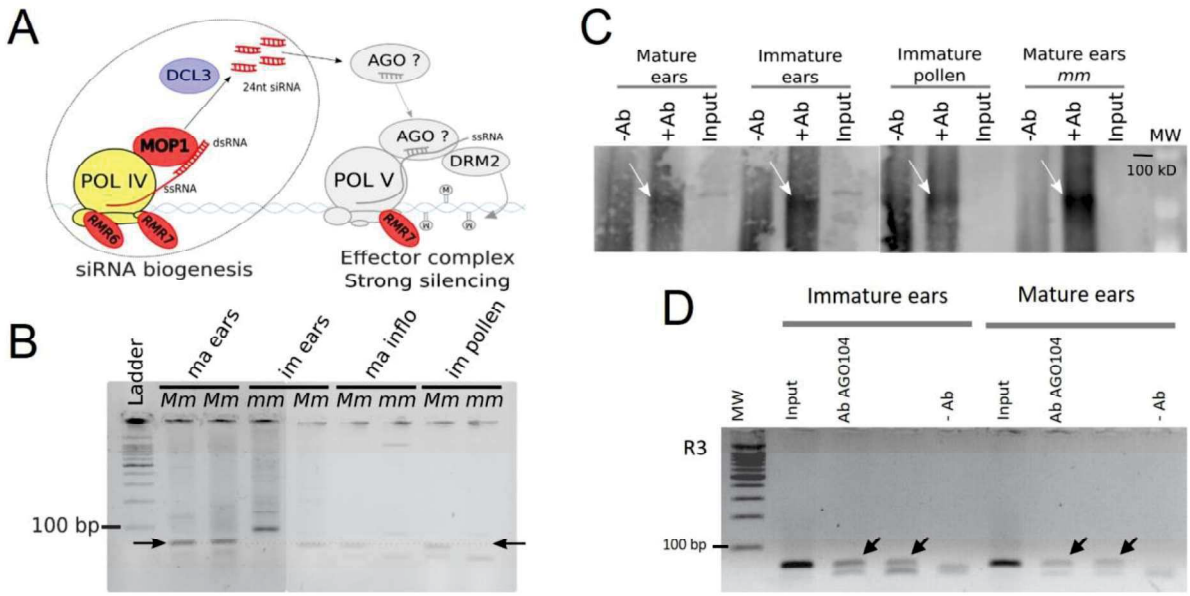



**Figure 1.** *b1TR* siRNAs and AGO104 interact in reproductive tissues. **A)** Identified (colored) and putative (based on homology with *A. thaliana* proteins; grey) RdDM members involved in small interfering RNAs biogenesis (left) and *de novo* methylation (right) in maize. RdDM proteins involved in paramutation are shown in red. Note that RMR7 is a subunit of both POL IV and POL V, therefore involved in both biogenesis and DNA methylation. **B)** Stem-loop PCR for R3 siRNAs in immature (im) and mature (ma) reproductive tissues of *Mm* and *mm* plants. Arrows indicate the 67-bp R3 expected band. **C)** Western blot of AGO104 in three reproductive tissues of MM and in mature *mm* ears. White arrows indicate the expected band. +Ab and -Ab are the IP samples treated with and without antibodies, respectively. Input is the sample that did not undergo IP. MW : molecular weight. **D)** Stem-loop PCR of siRNAs extracted from IPs of AGO104 in the *B'* genetic background. Input are small RNAs extracted directly from reproductive tissues. AbAGO104 are the small RNAs extracted from the IPs of AGO104. -Ab correspond to the mock immunoprecipitation samples (without Ab). Arrows indicate the 67-bp expected bands. MW: molecular weight.

To date, RdDM members found to affect paramutation in maize include MOP1 (Dorweiler et al., 2000; Alleman et al., 2006) and two REQUIRED TO MAINTAIN REPRESSION (RMR), RMR6/MOP3 that encodes the largest subunit of POL IV (Hollick et al., 2005; Erhard et al., 2009; Sloan et al., 2014) and RMR7/MOP2 that encodes a subunit shared between POL IV and POL V (Sidorenko et al., 2009; Stonaker et al., 2009). These proteins are essential to maintain paramutation states at *b1*, especially MOP1 as illustrated by the dark purple phenotype in *mop1* mutant progenies (Dorweiler et al., 2000). Interestingly, RMR7/MOP2 is required for both Pol IV and Pol V complex activity but its central role in POL IV prevents to observe its effect downstream of RdDM, in POL V. Therefore, no specific actor of the RdDM effector complex (later called effectors) has been identified yet in maize (Fig. **1A**).

*Zm*AGO104 has been proposed as a putative homolog of *At*AGO9 in maize (hereafter referred to as AGO104) (Singh et al., 2011). The goal of this work was to determine whether AGO104 is an effector of the RdDM complex, and whether it is involved in paramutation. Using immunolocalization and immunoprecipitation, we show that AGO104 is an effector of RdDM and has a similar function and localization to that of its homolog *At*AGO9 in *A. thaliana*. Next, sequencing of the small RNAs bound by AGO104 demonstrated that *b1TRs* of the *b1* enhancer region are RdDM target loci. Finally, we designed a reverse-genetics approach to functionally validate the role of AGO104 in paramutation by investigating paramutation-associated phenotypes in *mop1-1;ago104-5* stocks. Taken together, these data indicate that AGO104 is a member of the RdDM effector complex in maize and that it participates in paramutation at the *b1* locus. This research provides a deeper understanding of the molecular mechanisms underlying paramutation as well as new insights into the role of RdDM in maize.

## Results

### AGO104 is an effector of RdDM in reproductive tissues

To determine if AGO104 is an effector of RdDM we selected the *mop1-1/mop1-1* mutant (*mm*) that disrupts the RdDM pathway by decreasing the amounts of 24-nt siRNAs while it remains fully operational in heterozygous (*Mm*) plants (Dorweiler et al., 2000; Nobuta et al., 2008).

To validate our hypothesis that AGO104 is a functional homolog of AtAGO9, we first extracted total small RNAs from immature (at sporogenesis) and mature (at gametogenesis) ears, and mature pollen from both *Mm* and *mm* plants and investigated by stem-loop PCR the expression of 24-nt siRNAs previously identified as "RDR2-sensitive"(Nobuta et al., 2008; Jia et al., 2009). The detection of one of

these siRNAs, R3, in ears and pollen of *Mm* plants, but not in *mm* reproductive tissues (Fig. **1B**) confirmed that R3 siRNAs production is RdDM-dependent. Furthermore, we showed co-expression of AGO104 and R3 in mature and immature ears and pollen using Western blot analyses performed with a specific anti-AGO104 antibody (Singh et al., 2011) (Fig. **1C**). These results led us to perform AGO104 RNA immunoprecipitation (RNA-IP) on these reproductive tissues from *Mm* plants followed by small RNAs extraction and stem-loop PCR for R3. We detected a clear band of the expected size, indicating that AGO104 binds R3 siRNAs in reproductive tissues of maize (Fig. **1D**). This strongly suggests that AGO104 acts in RdDM, downstream of siRNA biogenesis (hence, downstream of MOP1).

Finally, we used Illumina sequencing of libraries prepared from the small RNAs previously recovered from AGO104 in immature ears of plants producing normal (B73 and *Mm*) and reduced (*mm*) amounts of 24-nt siRNAs. About two million cleaned reads were generated from each library and aligned onto the B73 reference genome (v5), and read counts were normalized using the transcripts per million (TPM) procedure. All genotypes displayed a similar chromosome-scale coverage using the TPM normalization procedure (Fig. **S1**). It is relevant to note that this method is used to normalize all backgrounds to a similar level of expression. We used this method to evaluate global chromosome coverage rather than differences in siRNAs expression levels. We evaluated the size of the reads and found that in plants producing reduced amounts of siRNA (*mm*) AGO104 binds more 21 and 22-nt small RNAs (Fig. **2A**) whereas in plants with regular abundance of siRNAs (B73 and *Mm*) it binds mostly 24-nt siRNAs. This change is probably caused by a decreased 24-nt siRNAs abundance in *mm* plants rather than by a change of AGO104 specificity (Nobuta et al., 2008). However, these results indicate that AGO104 binds 24-nt siRNAs in a non-mutant background, which strengthens our conclusion that AGO104 is an effector of RdDM.

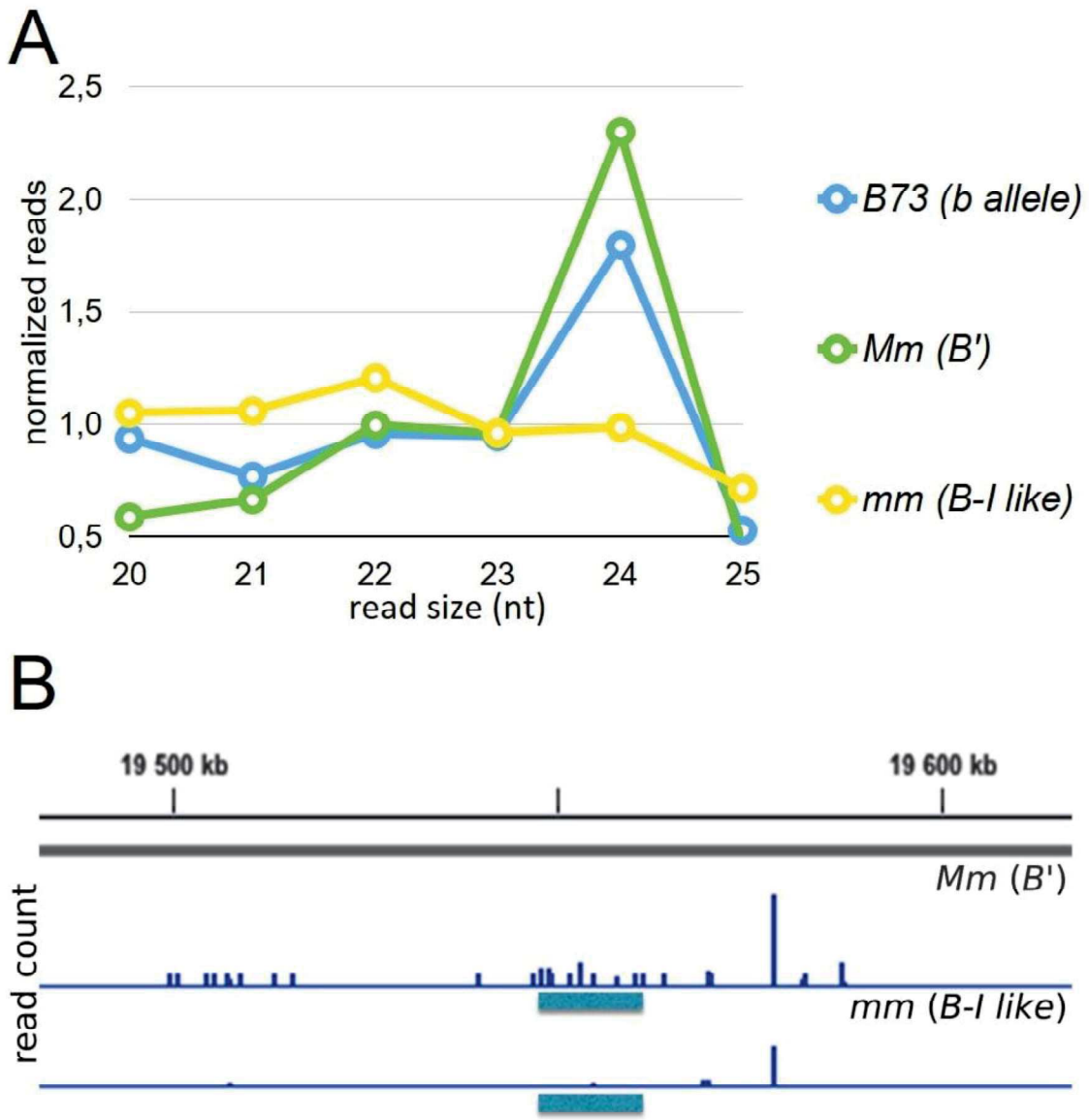



**Figure 2.** RIP-seq of AGO104-loaded small RNAs in mature ears of B73, *Mm* and *mm* individuals. **A)** Size distribution of reads normalized to 1. **B)** Distribution of 20 to 25-nt reads within the 100 kb region centered on the 853-bp *b1TR* region *(*blue horizontal box). Vertical blue bars indicate normalized read counts (Reads per million, RPM).

### AGO104 binds 24-nt siRNAs generated from *b1TR* sequences

As an effector of RdDM acting in reproductive tissues of maize, we wanted to determine whether AGO104 is a factor contributing to paramutation. It is worth noting that the *mop1-1* genetic stocks used in this research contain the *B'* allele (ie, its enhancer region harboring seven *b1TR* sequences) for which repressed state and paramutagenicity are both prevented by the MOP1 depletion causing a dark purple pigmentation. However, if transmitted in a MOP1 genetic background, the *B'* repressed state is

inherited and both light pigmentation and paramutagenicity are restored (Dorweiler et al., 2000). Therefore, the *B'* allele in *mop1* plants is an epiallele that is referred to as *B-I like* hereafter.

As a first step, we verified that AGO104 has the capacity to load 24-nt siRNAs associated with paramutation. To achieve this, we used the 20-to-25-nt reads from the above *Mm* and *mm* libraries (that are in the *B'* genetic background) and mapped them onto a composite segment assembled using the 100-kb region of the B73 reference genome centered on the *b1* enhancer region which we replaced by the *b1TR* repeats of *B'* genetic background (accession AF483657) (Stam et al., 2002a). The *b1* enhancer region in B73 was identified using sequence homology with the *b1TR* repeats. Interestingly, 20 to 25-nt small RNAs extracted from AGO104 in the *mm* mutant (*B-I like*, reduced amounts of 24-nt siRNAs) failed to map to the *b1TRs*, although AGO104 binds to 21 and 22-nt small RNAs as shown in Fig **2A**. However, 20 to 25-nt small RNAs extracted from AGO104 in *Mm* plants (B', producing normal amounts of 24-nt siRNAs) mapped correctly to the *b1TRs* region (Fig. **2B**). This data indicates that AGO104 from *Mm* (*B'*) binds 24-nt siRNAs that are produced from the *b1TR*.

Further evidence supporting this view arose from AGO104 immunolocalization experiments in young embryos, a tissue where paramutation is thought to be established (Hollick et al., 1995; Chandler et al., 2000; Haring et al., 2010). B73 embryos at three days after pollination (DAP) were dissected and treated using the AGO104 anti-body. As shown Fig. 2, AGO104 was expressed in the cytoplasm of embryonic cells, a cellular localization reminiscent to that of AGO9 in *A. thaliana* (Olmedo-Monfil et al., 2010). Altogether, our results support well the conclusion that *Ago104* in maize is an orthologue of *At*Ago9 and a strong candidate factor for paramutation.

## Setting up a reverse genetics approach for testing the role of AGO104 in paramutation at the *b1* locus

While *B'* paramutagenic alleles are highly stable, *B-I* paramutable alleles are unstable and can spontaneously change into *B'* with a wide range of frequencies (from 0.1 up to > 50%) depending on the genetic background (Chandler et al., 2000; Sidorenko et al., 2009). To avoid drawbacks resulting from instability of naïve *B-I* alleles and to ensure that the anthocyanin biosynthesis pathway is functional in the genetic background used, we used the properties acquired by *B'* alleles when introduced into *mop1* plants. Both paramutagenicity and phenotypes of *Mm* (*B'*) and *mm* (*B-I like*) plants and their behavior upon crossing with neutral b alleles was extensively studied (Dorweiler et al., 2000; Alleman et al., 2006). Although *B-I like* plants display a dark purple phenotype, both *B-I like* /*b*

and *B-I like* /*B'* progenies are lightly pigmented plants as the *B'* phenotype is not heritably altered by *mop1* mutations. This property of *B-I like* alleles was used to create a reverse genetic population.

We selected *ago104-5,* a Mutator-induced allele previously characterized as a dominant knockout allele creating defects during female meiosis and apomixis-like phenotypes (Singh et al., 2011). This mutation was introduced in the B73 inbred background that carries a neutral *b* allele (Chandler, 2004). Since previous works in *A. thaliana* reported that reduced levels of small RNAs alter AGO proteins expression (Havecker et al., 2010), we first verified whether AGO104 expression was altered in *mm* reproductive tissues. Using Western blot, AGO104 was detected in *mm* reproductive tissues suggesting that, contrary to that observed in *A. thaliana*, reduced levels of small RNAs did not alter the presence of the AGO104 protein in maize tissues (Fig. **1c**). The *ago104-5* mutation and the *mop1-1* mutation were then combined to generate genotypes of interest for investigating the contribution of AGO104 to paramutation and for controlling our experiment.

We first crossed *mm* plants (dark purple, *B-I like*) with *ago104-5* (*aa*) plants (green, neutral *b*) (Fig. **3** – cross 1). We evaluated stem pigmentation of the resulting F1 progeny (n=14) that consisted of double heterozygous (*Mm;Aa*). We then backcrossed F1s (*Mm;Aa*) to the *mm* mutant (Fig. **3** – cross 2) which generated progenies either functional (*Mm*) or deficient (*mm*) for MOP1. The *Mm* population was conserved for phenotype evaluation while the *mm* population was used as control plants for functional anthocyanin biosynthesis pathway, as dark pigmentation is expected regardless of their genotype at the *Ago104* locus (MOP1 acts upstream in the RdDM pathway). We determined the genotype of the progeny at the *Mop1* and *Ago104* loci and evaluated stem pigmentation at 46 and 56 days post-seeding (dps). Finally, to control environmental effects, we also evaluated stem pigmentation of both *Mm* (*B')* and *mm* (*B-I like*) plants derived from stocks segregating the *mop1-1* allele only. Also note that in previous crosses between *Mm* plants and B73 inbred (*b* allele), no significant effect on plant pigmentation was reported (Barber et al., 2012; Madzima et al., 2014), therefore indicating that the B73 genome does not harbor factors affecting pigmentation at the *b1* locus.

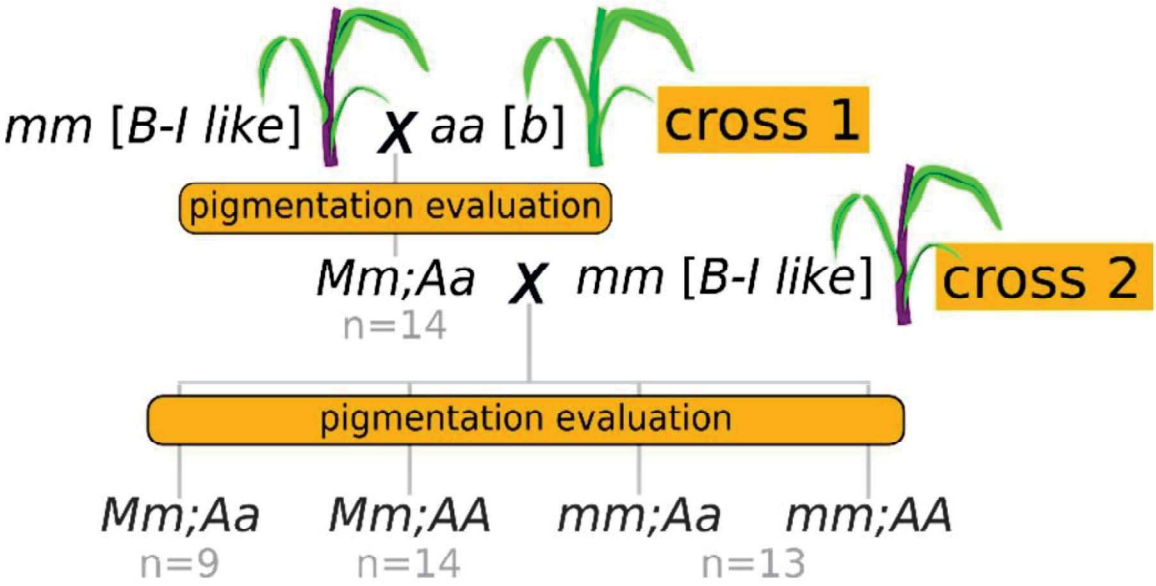



**Figure 3.** Crossing scheme used for a reverse genetic screen designed to investigate AGO104 contribution to paramutation. Allele designation is as follows: *M* : *Mop1; m: mop1-1; A*: *Ago104; a* : *ago104-5; b*: neutral *b1* allele; *B-I like*: *mm* plant. Purple and green stem drawings represent respectively dark purple and green stem phenotypes. Alleles at the *b1* locus are indicated in square brackets.

### *ago104-5* mutation disrupts paramutation at the *b1* locus

As expected from previous works for control plants (segregating the *mop1-1* allele only) (Dorweiler et al., 2000; Alleman et al., 2006), all *mm* plants (*B-I like*, n>25) were dark purple at 46 and 56 days post seeding (dps), while all *Mm* plants (*B'*, n>25) were lightly pigmented (Fig. **4A-C-D**). In cross 2 progeny, all *mm* progeny (*B-I like*, n=13) displayed the same dark purple phenotype as seen in control *mm* plants and regardless of the *ago104* genotype (Fig. **4B**). This was expected since MOP1 acts upstream of AGO104 in RdDM and indicates that AGO104 unlikely contributes to paramutation through another, however unknown, mechanism. Interestingly, F1 plants (n=14) displayed a new phenotype with intermediate levels of pigmentation, seemingly a partially reverted paramutation phenotype. Furthermore, *Mm* plants from cross 2 (n=23) also displayed this new phenotype and were tested for pigmentation dynamics which showed an increase between 46 and 56 dps. At 46 dps, 30% (n=7) showed a typical light phenotype, while the remaining plants (n=16) exhibited a partially reverted paramutation phenotype with intermediate levels of pigmentation and nodes lacking pigmentation (Fig. **4B-E**). Pigmentation turned darker over time in all *Mm* progenies, none of which exhibiting at 56 dps lightly pigmented stem. 52% of the progeny (n=12) reached an intensity similar to that conferred by the *B-I like* allele in *mop1-1* mutants, and intermediate levels in the remaining plants (n=11) (Fig.

**4B**). Notably, however, pigmentation around nodes remained weak throughout development, a phenotype not observed nor previously reported during *mm* plants development.

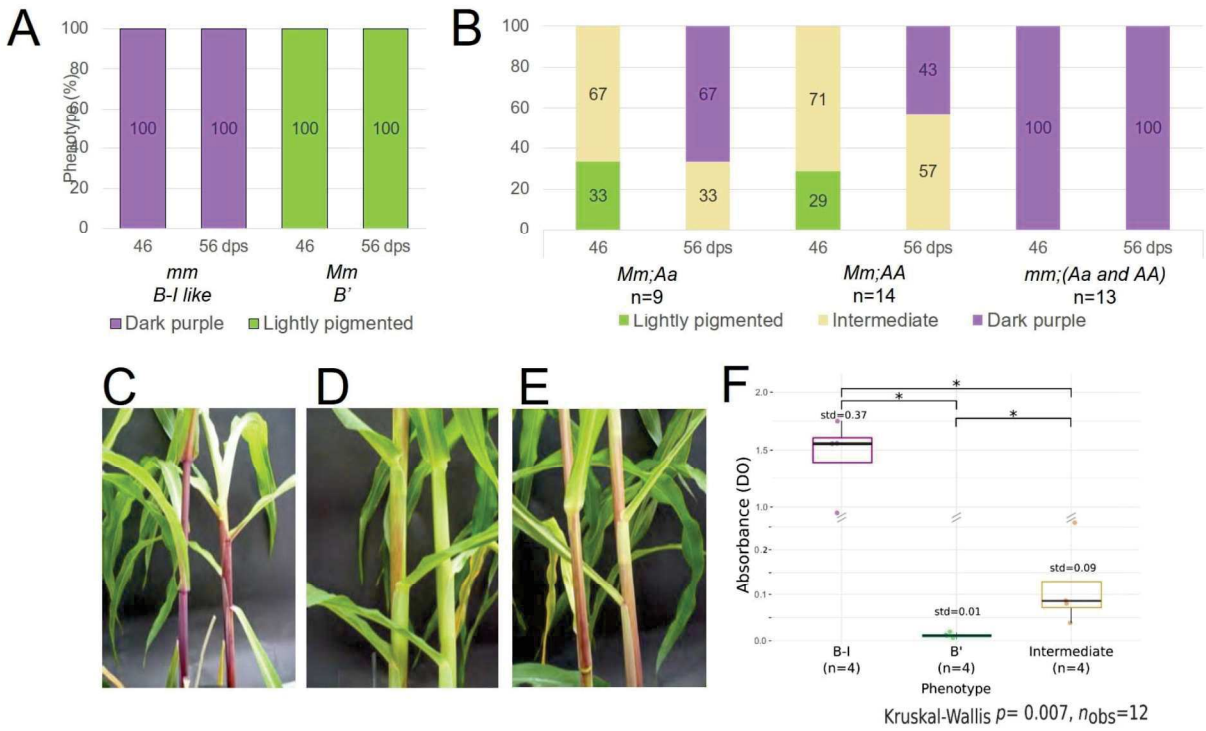



**Figure 4.** Pigmentation phenotypes observed at 46 and 56 days post seedling (dps) in **A**) control plants (with n>25 for each control) and **B**) the cross 2 progenies. **C-E)** Illustration of stem pigmentation phenotypes observed at 46 dps in **c**) control *mm* plants and *mm* cross 2 progenies (dark purple); **D**) *MM* and *Mm* control plants (light pigmentation), and; **E**) in *Mm* cross2 progenies (intermediate pigmentation). **F)** Absorbance at 550 nm of anthocyanins extracted from 1 g of stem tissue from 56 dps plants. *B-I*: *mm* mutant (*B-I like*); *B'* : *Mm* plants; Intermediate: *Mm;Aa* plants. std is the standard deviation. Y axis is represented between 0< DO <0.25 (scale = .05) and 1< DO <2 (scale = .25).

To further our visual observations, we quantified anthocyanins by spectrophotometry in extracts obtained from stem tissues collected at 56 dps from plants with intermediate levels of stem pigmentation (*Mm* progeny from cross 2) and, as control for pigmentation, dark purple plants (*mm*, *B-I like*) and lightly pigmented plants (*Mm*, *B'*). Both Kruskal-Wallis test (p = .007) and multiple pairwise comparison test (p = .029) indicated a significant difference in pigment quantifications between the three classes (Fig. **4F**). These results suggest that a gradual release of *B'* silencing allowed increasing anthocyanin production in all progenies although it never achieves similar pigmentation level to that of *B-I like* in the *mop1-1* condition.

As *Mm* progenies used here varied for the *Ago104* genotype (*AA* : *Aa* followed the expected 1:1 ratio; Chi2 value 1.09, $p < .05$), we wanted to determine whether this condition could be associated with differences in pigmentation intensity. Fisher exact test revealed no significant association between the *Ago104* genotypes and pigmentation in *Mm* progenies (value 0.69, $p < .05$; see Fig **4** for category numbers). In other words, the intermediate phenotype identified in the *Mm* progeny from cross 2 happens with similar proportions in *Aa* and *AA* plants. This suggests that either a parental effect or a heritable release of silencing of *B'* alleles are mediated by an AGO104 deficiency in the progeny of cross 1. Therefore, mutation of *Ago104* alters the paramutation state of *B-I like* epialleles when transmitted through meiosis.

## Discussion

Given the central role of *At*AGO4 in RdDM in *A. thaliana*, we first considered its two closest homologs in maize, *Zm*AGO105 and *Zm*AGO119 for our study. However, high sequence similarity between the two maize sequences (Singh et al., 2011) rendered difficult to generate specific antibodies, and makes it likely that these genes can complement their mutant homolog. On the other hand, another ARGONAUTE protein, *At*AGO9, plays a crucial role in RdDM in *A. thaliana* although it does not fully complement *At*AGO4 when regulated by AGO4's promoter (Havecker et al., 2010). *At*AGO9's expression in reproductive tissues is of particular interest with regards to the establishment of paramutation. Our results show that *Zm*AGO104, the maize putative homolog of AtAGO9, is cytoplasmic (Fig. **S2**), a location similar to that of AtAGO9 (Olmedo-Monfil et al., 2010). Based on these results and on sequence similarities previously reported (Singh et al., 2011), we argue that maize *Ago104* is a functional ortholog of *AtAGO9*, an assumption strongly supported by our smRNAseq from AGO104 IP analysis indicating AGO104 preferentially recruits 24-nt small RNAs, including those generated from *b1TR*s and involved in paramutation (Havecker et al., 2010).

Despite functional similarities, AGO104 and *At*AGO9 might be differently regulated. In the *rdr2* mutant in *A. thaliana*, the levels of 24-nt siRNAs are strongly reduced which causes instability of the AGO4 group proteins (Havecker et al., 2010). Decrease in 24-nt siRNAs production in the *mop1-1* mutant in maize does not seem to influence the AGO104 stability in mature ears as shown by our Western blot experiments (Fig. **1C**). This suggests that either the degradation mechanisms are different in maize or there is a pathway capable of rescuing/stabilizing AGO104 independently of MOP1-dependent siRNAs. As previously shown, several "alternative" RdDM pathways enable the synthesis of 24-nt siRNAs without the involvement of RDR2/MOP1 in *A. thaliana* (Cuerda-gil and Slotkin, 2016) and maize

(Nobuta et al., 2008; Wang et al., 2017). However, the production of 24-nt small RNAs in the *mop1-1* mutant is partially replaced by 22-nt small RNAs (Nobuta et al., 2008). This supports our results by which AGO104 proteins in *mop1-1* mutant did not carry 24-nt siRNAs, and loaded preferentially 22-nt RNAs (Fig. **2A**). A possible explanation for this might be that the 22-nt small RNAs in *mop1-1* mutant contribute to rescue AGO104, but they do not mediate paramutation at the *b1* locus.

Our reverse genetic screening performed on *ago104-5* mutants helps in understanding AGO104 involvement in paramutation. Paramutation at the *b1* locus involves the *B-I* and *B'* alleles, respectively associated with the typical intense and light plant pigmentation (Coe, 1959). Here, our reverse genetics approach combining *ago104-5* and *mop1-1* mutations unveiled an intermediate pigmentation phenotype that turns darker over time (Fig. **4A**). However, although pigmentation in these plants seems to reach that of *mm* plants at 56 dps, quantification using spectrophotometry showed that *mm* plants produce higher levels of anthocyanins. Previous description of the *mop2* mutant also reported an evolution of pigmentation over time that never reaches *mm* plants levels (Sidorenko et al., 2009). Both *mop2* and *mop1* mutants alter siRNAs production and potentially have effects beyond those resulting from RdDM downregulation (reviewed in (Cuerda-gil and Slotkin, 2016)). In contrast, mutations of AGO104 perturb RdDM targeting but not 24-nt siRNA production nor their possible contribution to paramutation through other regulatory mechanisms (reviewed in (Vaucheret, 2008)).

Interestingly, all *Mm* plants (F1s and cross 2 progeny) displayed the same intermediate phenotype, demonstrating that AGO104 is an effector of paramutation and suggesting that the *ago104-5* mutation does not allow a complete reversion to the *B-I* dark purple phenotype. Other AGO proteins, such as AGO105 and AGO119, may complement AGO104 depletion and, thus, ensure some silencing of the *b1* gene, preventing the full reversion to the *B-I* phenotype. Furthermore, both F1 plants and their *Mm;AA* progeny displayed an intermediate phenotype, suggesting that the *ago104-5* mutation alters the *B'* paramutation state through meiosis, and disrupts the heritability of paramutation at the *b1* locus. Such reversion of paramutation was previously described at the *Pl'* allele in the *mop1-1* mutant (Dorweiler et al., 2000).

Consistent with our results of RNA-IP and immunolocalization, previous studies have demonstrated that AGO104 is located exclusively in reproductive tissues (i.e. female and male meiocytes, egg cells, and embryos) (Singh et al., 2011), where paramutation is at least partly established (Hollick et al., 1995; Chandler et al., 2000; Haring et al., 2010). Interestingly, *b1* is expressed in somatic tissues only (Chandler, 2004; Arteaga-Vazquez et al., 2010), where maintenance of paramutation takes place and

where AGO104 is not expressed. Hence, AGO104 is probably involved in the establishment rather than the maintenance of paramutation. Interestingly, we observed green nodes in *Mm* plants in the progeny of cross 2 with intermediate pigmentation. No previous research was conducted to study the specific behaviour of meristematic tissue in paramutation, but genes involved in regulation of maize development are themselves regulated by regulators of paramutation like MOP1 and RMR6/MOP2 (Dorweiler et al., 2000; Hollick et al., 2005; Parkinson et al., 2007). Therefore, it is possible that meristematic tissues possess backup mechanisms to regulate their development and, at the same time, can establish paramutation contrary to somatic tissues.

In this study, we identified a new effector of RdDM in maize using Western blot and RNA Immunoprecipitation. We also confirmed that AGO104 binds paramutation-associated siRNAs by sequencing small RNAs loaded onto AGO104, and our reverse genetic approach validated the functional role of AGO104 in paramutation at the *b1* locus. AGO104 is involved in the establishment of paramutation in the reproductive tissues of maize, most likely through its function in the effector complex of the RdDM pathway. While other AGO proteins might play similar functions as AGO104 in RdDM and paramutation, our findings shed new light on the mechanisms mediating both the establishment and the transmission of paramutation in maize.

## Materials and methods

### Plant material

The B73 inbred line was provided by the Maize Genetics Cooperation Stock Center (University of Illinois, Urbana/Champaign, USA). The Trait Utility System for Corn (TUSC) at Pioneer Hi-Breed (Johnston, Illinois, USA) provided *ago104-5* stocks and V.L. Chandler (University of Arizona, Tucson, AZ, USA) provided the *mop1-1* mutant in the *B'* genetic background. Plants were grown in a greenhouse at the French National Research Institute for Sustainable Development in Montpellier, France, with 14 hours day light (26°C during the day, 20°C at night). For all these plants, inflorescences were partially dissected to evaluate pollen developmental stages with a Zeiss confocal microscope. We snap froze and stored at -80°C both inflorescences collected at sporogenesis and gametogenesis stages (respectively, immature and mature inflorescences), and pollen during sporogenesis (immature pollen). Ears at sporogenesis (immature ears) and at gametogenesis (mature ears) were selected based on their length (3 to 5 cm of length for immature ears and > 5 cm for mature ears) and the presence of silks, and were immediately snap frozen in liquid nitrogen and stored at -80°C.

## Genotyping

Total genomic DNA was extracted from seedling tissues using a standard CTAB procedure. After quality check for DNA integrity and quality, DNA concentration was quantified using a NanoDrop spectrophotometer. Genotyping was performed by PCR using 20 µl reactions containing 200 ng DNA, 1 µL of 10 µM of forward and reverse primers (see Table S3) and Quick-Load Taq 2X Master Mix (NewEngland Biolabs). For amplifications, samples preparation were denatured for 3 mn at 95°C and subjected to 35 cycles as follows: 15 s at 95°C for denaturation, 15 s at 55°C for annealing and, 60 s and 165 s extension at 68°C for *mop1-1* and *ago104-5*, respectively. Amplification products were loaded in 1.5% agarose gels, electrophorized at 100 V for 20 min and visualized by ethidium bromide staining.

## Immunolocalization

Two fertilized ovaries from B73 plants were collected 3 days after pollination (DAP) and sliced using a Vibratom (Leica VT1000E) into 200 to 225 µm sections. They were immersed 2 h in a fixative solution (4% paraformaldehyde, PBS 1X, 1% Tween 20, 0.1 mM PMSF) and washed 3 times in PBS (Phosphate Buffered Saline). Samples were then digested for 15 min at room temperature using an enzymatic solution (1% driselase, 0.5% cellulase, 1% pectolyase, 1% BSA, all from Sigma-Aldrich), and washed 3 times in PBS. Samples were left 1 h in permeabilizing solution (PBS 1X, 2% Tween 20, 1% BSA) on ice, washed 3 times in PBS and incubated overnight at 4°C in an AGO104 primary antibody (S1 Table) diluted 1:50 inb PBS. Samples were left 8 h in a PBS 1X-0,2% Tween 20 washing solution with solution renewal every 2 h. They were incubated overnight with Alexa Fluor 488 secondary antibody (1:200 in PBS) and left 6 h in washing solution. They were then incubated 1 h in DAPI, rinsed with PBS 1X, and mounted in ProLong Antifade Reagent (Invitrogen). Slides were sealed with nail polish and stored at -20°C. Observations were made using LEICA SPE with 405 nm (DAPI) and 488 nm (Alexa Fluor 488) excitation.

## Small-RNA Immunoprecipitation

Protocols were adapted from (Havecker et al., 2010) using two biological replicates per genotype. Tissues were grinded with liquid nitrogen and a Dounce homogenizer. Resulting powder was placed in a Falcon tube with 3 volumes of extraction buffer (20 mM Tris HCL pH 7.5, 5 mM $MgCl_2$, 300 mM NaCl, 0.1% NP-40, 5 mM DTT, 1% protease inhibitor (Roche Tablet), 100 units/mL RNase-OUT (Invitrogen)). Samples were vortexed, kept on ice 30 min with continuous shaking, and centrifuged for 20 min at 4°C (4000 rpm). Supernatant was filtered through a 0.45 µm filter into a new Falcon tube and 1 mL was

aliquoted and stored at -20°C as a pre-experiment input sample. We generated 2 mL aliquots from the remaining samples and added 5 µg of antibodies per gram of tissue. 20 µL of prepared Dynabeads (Life technologies) magnetic beads (ie, washed 3 times in wash buffer (20 mM Tris HCL pH 7.5, 5 mM $MgCl_2$, 300 mM NaCl, 0.1% NP-40, 1% protease inhibitor, 100 units/mL RNase OUT), were added to each sample. After 2 h incubation on a rotation wheel at 4°C, the beads were washed 3 times and resuspended in 500 µL of washing buffer. 100 µL was aliquoted and stored at -20°C for Western blot control experiments. The washing buffer was then discarded and replaced by 250 µL of elution buffer prepared according to (Terzi and Simpson, 2009) (100 mM NaHCO3, 1% SDS, 100 units/mL RNase OUT (Invitrogen) in 0.1% DEPC water), and the tubes were incubated 15 min at 65°C with agitation. Supernatant was transferred to fresh tubes and elution was repeated once. The two eluates were finally combined. Samples were treated with 0.08 µg/µL proteinase K for 15 min at 50°C. RNA was extracted following the recommendations from Applied Biosystems for TRI Reagent® Solution, starting by adding 1.2 mL of TRI Reagent to the samples.

### Control for RNA-IP: Western blot

The protocols were adapted from (Martínez-García et al., 1999). The control samples from the RNA-IP were added in Laemmli 4X buffer (250 mM Tris pH 6.8, 4% SDS, 20% glycerol, 10 mM DTT, Bromophenol Blue) and incubated 15 min at 90°C. Samples (15 µl) were migrated at 180 V for 45 min in 1X migration buffer (25 mM Tris base, 190 mM glycine, 0.1% SDS) together with PageRuller Plus Ladder (ThermoFisher Scientific) for protein size standards. Transfer onto nitrocellulose membrane (Amersham) was achieved in transfer buffer (migration buffer 1X, 20% ethanol), at 125 V for 2.5 h. Membranes were then rinsed in PBST and left 1 h in PBST supplemented with 5% non-fat milk (PBST-milk solution). After renewing PBST-milk, 1:200 AGO104 antibody (S1 Table) was added and left overnight with agitation. After 4 x 5 min washes in PBST-milk with agitation, PBST-milk was added with 1:2500 HRP antibody (Invitrogen) and membranes were incubated 2 h with agitation. Membranes were washed 4 times with PBST and prepared for analysis with a Typhoon 9400 as recommended by the ECL plus Western Blotting detection system (Amersham). Finally, they were washed in PBST, stained 30 min in Ponceau S Solution (Sigma-Aldrich) with agitation, and rinsed before visualizing proteins.

### Stem loop PCR

Small RNAs extracted from RNA-IPs were treated with DNase to remove potential DNA contamination using the TURBO DNA-free kit (AM1907, Ambion Life technologies). DNA-free samples, 50 µM of stem-loop primers (listed in S2 Table), 10 mM of dNTP and nuclease-free water were mixed to reach a final

volume of 13 µL. Stem-loop reverse transcription was performed following the recommendations from (Varkonyi-Gasic et al., 2007) and the resulting double-stranded cDNAs were used for PCR. 1 µL of cDNA was mixed with Red Taq 2x (Promega), and 0,25 µM of universal reverse primer (complementary to the stem loop one) and a specific forward primer designed to match the *b1TR* siRNAs. 20 µL reactions were denatured for 2 min at 94°C, and went through 40 cycles of 15 s at 94°C and 1 min at 60°C. Product visualization was performed by electrophoresis into 2% agarose gels (Lonza) in TBE 0.5X supplemented with 0.5 µg/mL BET for 40 min at 100 V. To verify cDNAs derived from *b1TR* siRNAs, amplified products were recovered from the gel using the QIAquick gel extraction kit (QIAGEN) and cloned in DH5α competent cells (Invitrogen) using the pGEM-T Easy Vector Systems protocol (Promega) and an LB-ampicillin selective medium. Colonies were genotyped using the T7/SP6 primers (Promega). Plasmids from selected colonies were isolated using the QIAprep Spin Miniprep Kit (QIAGEN) and sequenced (Beckman Coulter Genomics, Inc., UK).

### Small RNA sequencing

Small RNAs extracted from RNA-IPs were migrated on a 1.5% agarose gel and recovered from the corresponding bands using the Monarch DNA Gel Extraction kit (NEB #T1020 New England Biolab). RNA samples were used to prepare libraries using the NEBNext Multiplex Small RNA Library Prep Set (NEB #E7300S New England Biolab). The final PCR enrichment was performed using 15 cycles. Samples were quantified with Qubit and Agilent Bioanalyzer using the DNA high-sensitivity assays and sequenced on a NextSeq550 machine at the CSHL Genome Center.

### Small RNA seq analysis

Raw reads were cleaned up using Trimmomatic (Version 0.38) with the following parameters 2:30:5 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:15 MAXLEN:35. Cleaned reads were first aligned to the maize reference genome B73 version 5 using Bowtie 1 (Version 1.2.2) with the following options --best -k 2 -5 4 -p 10. Mapped reads coverage into 0.5 Mb genome windows was generated using the coverage utility of the Bedtools suite (Quinlan and Hall, 2010). For a better resolution, reads were also aligned to the *b1TRs* and their 100 kb flanks using Bowtie 1 (Version 1.2.2) with the following options -m 7 - q --strata --best -v 2. They were intersected into 50 bp genome windows using Bedtools coverage.

## Quantification of plant pigmentation

Adapting a protocol from (Chen et al., 2012), we collected and froze stem tissue from the seventh leaf of plants at 56 days post-seeding (dps) with light stem pigmentation (heterozygous *mop1-1*), intermediate stem pigmentation (plants from cross 2) and dark purple stem pigmentation (homozygous *mop1-1*). 1 g of tissues was grinded in liquid nitrogen and incubated in 30 mL of methanol:water (70:30) for 24 h at 4°C. Tubes were then centrifugated at 5,000 g for 30 min, and the supernatant was collected and centrifugated for 10 more minutes. Supernatant was then assessed for absorbance at 550 nm. Differences in absorbance between the 3 phenotypes was tested using the Kruskal-Wallis test and a multiple pairwise comparison test. For plants at 46 dps, light, intermediate and dark purple stem pigmentation was estimated visually, using the recognizable green area around nodes (which is not observed in *mop1-1*).


## Funding

D.G. received support from the H2020 MCA (REP-658900-2) and *Agence Nationale de la Recherche* grants REMETH (ANR-15-CE12-0012-03). D.G., O.L. and R.A.M. received support from the CHROMOBREED (ANR-18-CE92-0041). M.A.A.V. received support from the *Jeunes équipes associées à l'IRD (JEAI)* program (EPIMAIZE), Agropolis Foundation, CONACYT (158550 & A1-S-38383) and the Royal Society Newton Advanced Fellowship (NA150181). O.O.L. was the recipient of a graduate scholarship from CONACYT and from the *Bourses d'échanges scientifiques et technologiques IRD (BEST)* program. R.A.M. is supported by the Howard Hughes Medical Institute, and grants from the National Science Foundation. The authors acknowledge assistance from the Cold Spring Harbor Laboratory Shared Resources, which are funded in part by the Cancer Center Support Grant (5PP30CA045508).


## Data Availability

Small RNA sequencing data were deposited in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) under the accession number GSE172479.

**Author Contributions:** J.A., F.B., O.L., D.G. designed the research; J.A, F.B., O.O-L, D.G. performed the research; J.A., F.B., O.L., D.G. analysed the data; R.A.M., D.G., M.A.A-V. acquired funding; all authors wrote the paper.

**Competing Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References


**Alleman M, Sidorenko L, McGinnis K, Seshadri V, Dorweiler JE, White J, Sikkink K, Chandler VL** (2006) An RNA-dependent RNA polymerase is required for paramutation in maize. Nature **442**: 295–298

**Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K, Green PJ, Meyers BC, Chandler VL** (2010) RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proc Natl Acad Sci **107**: 12986–12991

**Arteaga-Vazquez MA, Chandler VL** (2010) Paramutation in maize: RNA mediated trans-generational gene silencing. Curr Opin Genet Dev **20**: 156–163

**Barber WT, Zhang W, Win H, Varala KK, Dorweiler JE, Hudson ME, Moose SP** (2012) Repeat associated small RNAs vary among parents and following hybridization in maize. Proc Natl Acad Sci U S A **109**: 10444–10449

**Barbour JER, Liao IT, Stonaker JL, Lim JP, Lee CC, Parkinson SE, Kermicle J, Simon SA, Meyers BC, Williams-Carrier R, et al** (2012) Required to maintain repression2 is a novel protein that facilitates locus-specific paramutation in maize. Plant Cell **24**: 1761–1775

**Belele CL, Sidorenko L, Stam M, Bader R, Arteaga-Vazquez MA, Chandler VL** (2013) Specific Tandem Repeats Are Sufficient for Paramutation-Induced Trans-Generational Silencing. PLoS Genet. doi: 10.1371/journal.pgen.1003773

**Brink RA** (1973) Paramutation. Annu Rev Genet **7**: 129–152

**Chandler VL** (2004) Poetry of b1 paramutation: cis- and trans-chromatin communication. Cold Spring Harb. Symp. Quant. Biol. pp 355–361

**Chandler VL, Eggleston WB, Dorweiler JE** (2000) Paramutation in maize. Plant Mol Biol **43**: 121–145

**Chen S, Fang L, Xi H, Guan L, Fang J, Liu Y, Wu B, Li S** (2012) Simultaneous qualitative assessment and quantitative analysis of flavonoids in various tissues of lotus (Nelumbo nucifera) using high performance liquid chromatography coupled with triple quad mass spectrometry. Anal Chim Acta **724**: 127–135

**Coe EH** (1959) A Regular and Continuing Converstion-Type Phenomenon at the B Locus in Maize. Proc Natl Acad Sci 828–832

**Cuerda-gil D, Slotkin RK** (2016) Non-canonical RNA-directed DNA methylation. Nat Plants. doi: 10.1038/NPLANTS.2016.163

**Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL** (2000) *mediator of paramutation1* Is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci. Plant Cell **12**: 2101–2118

**Erhard KF, Parkinson SE, Gross SM, Barbour JER, Lim JP, Hollick JB** (2013) Maize RNA polymerase IV defines trans-generational epigenetic variation. Plant Cell **25**: 808–819

**Erhard KF, Stonaker JL, Parkinson SE, Lim JP, Hale CJ, Hollick JB** (2009) RNA polymerase IV functions in paramutation in Zea mays. Science (80- ) **323**: 1201–1205

**Giacopelli BJ, Hollick JB** (2015) Trans-Homolog interactions facilitating paramutation in maize. Plant Physiol **168**: 1226–1236

**Haring M, Bader R, Louwers M, Schwabe A, Van Driel R, Stam M** (2010) The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. Plant J **63**: 366–378

**Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH, Baulcombe DC** (2010) The arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. Plant Cell **22**: 321–334

**Hollick JB** (2017) Paramutation and related phenomena in diverse species. Nat Rev Genet **18**: 5–23

**Hollick JB, Kermicle JL, Parkinson SE** (2005) Rmr6 maintains meiotic inheritance of paramutant states in Zea mays. Genetics **171**: 725–740

**Hollick JB, Patterson GI, Coe EH, Cone KC, Chandler VL** (1995) Allelic interactions heritably alter the activity of a metastable maize pl allele. Genetics **141**: 709–719

**Hövel I, Pearson NA, Stam M** (2015) Cis-acting determinants of paramutation. Semin Cell Dev Biol **44**: 22–32

**Jia Y, Lisch DR, Ohtsu K, Scanlon MJ, Nettleton D, Schnable PS** (2009) Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. PLoS Genet. doi: 10.1371/journal.pgen.1000737

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

**Madzima TF, Huang J, McGinnis KM** (2014) Chromatin structure and gene expression changes associated with loss of MOP1 activity in Zea Mays. Epigenetics **9**: 1047–1059

**Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM** (2006) The evolution and diversification of Dicers in plants. FEBS Lett **580**: 2442–2450

**Martínez-García JF, Monte E, Quail PH** (1999) A simple, rapid and quantitative method for preparing Arabidopsis protein extracts for immunoblot analysis. Plant J **20**: 251–257

**Matzke MA, Kanno T, Matzke AJM** (2015) RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. Annu Rev Plant Biol **66**: 243–267

**Matzke MA, Mosher RA** (2014) RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. Nat Rev Genet **15**: 394–408

**Nobuta K, Lu C, Shrivastava R, Pillay M, De Paoli E, Accerbi M, Arteaga-Vazquez M, Sidorenko L, Jeong DH, Yen Y, et al** (2008) Distinct size distribution of endogenous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. Proc Natl Acad Sci U S A **105**: 14958–14963

**Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Arévalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP** (2010) Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature **464**: 628–632

**Parent JS, Cahn J, Herridge RP, Grimanelli D, Martienssen RA** (2021) Small RNAs guide histone methylation in Arabidopsis embryos. Genes Dev **35**: 841–846

**Parkinson SE, Gross SM, Hollick JB** (2007) Maize sex determination and abaxial leaf fates are canalized by a factor that maintains repressed epigenetic states. Dev Biol **308**: 462–473

**Quinlan AR, Hall IM** (2010) BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841–842

**Sidorenko L, Dorweiler JE, Cigan AM, Arteaga-Vazquez M, Vyas M, Kermicle J, Jurcin D, Brzeski J, Cai Y, Chandler VL** (2009) A dominant mutation in mediator of paramutation2, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. PLoS Genet. doi: 10.1371/journal.pgen.1000725

**Singh M, Goel S, Meeley R, Dantec C, Parrinello H, Michaud C, Leblanc O, Grimanelli D** (2011) Production of Viable Gametes without Meiosis in Maize Deficient for an ARGONAUTE Protein. Plant Cell **23**: 443–458

**Sloan AE, Sidorenko L, McGinnis KM** (2014) Diverse gene-silencing mechanisms with distinct requirements for RNA polymerase subunits in Zea mays. Genetics **198**: 1031–1042

**Stam M, Belele C, Dorweiler JE, Chandler VL** (2002a) Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev **16**: 1906–1918

**Stam M, Belele C, Ramakrishna W, Dorweiler JE, Bennetzen JL, Chandler VL** (2002b) The regulatory regions required for B' paramutation and expression are located far upstream of the maize b1 transcribed sequences. Genetics 162(2) 917-930

**Stonaker JL, Lim JP, Erhard KF, Hollick JB** (2009) Diversity of Pol IV function is defined by mutations at the maize rmr7 locus. PLoS Genet. doi: 10.1371/journal.pgen.1000706

**Terzi LC, Simpson GG** (2009) Arabidopsis RNA immunoprecipitation. Plant J **59**: 163–168

**Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP** (2007) Protocol: A highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Methods **3**: 12

**Vaucheret H** (2008) Plant ARGONAUTES. Trends Plant Sci **13**: 350–358

**Wang P-H, Wittmeyer KT, Lee T, Meyers BC, Chopra S** (2017) Overlapping RdDM and non-RdDM mechanisms work together to maintain somatic repression of a paramutagenic epiallele of maize pericarp color1. PLoS One **12**: e0187157

**Zhang H, Lang Z, Zhu JK** (2018) Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol **19**: 489–506

**Zhang Z, Teotia S, Tang J, Tang G** (2019) Perspectives on microRNAs and phased small interfering RNAs in maize (Zea mays L.): Functions and big impact on agronomic traits enhancement. Plants. doi: 10.3390/plants8060170

## Supplemental data

The following Supporting Information is available for this article:

**Fig. S1** siRNA chromosome coverage on the B73 reference genome (version 5).

**Fig. S2** Fluorescence of AGO104 in nuclei of B73 embryonic cells 3 days after pollination (DAP).

**Fig S3.** Complete gels presented in Fig. **1**.

**Table S1** AGO104 antibody characteristics.

**Table S2** Primer sequences used for siRNA stem loop PCR.

**Table S3** Primer sequences used for genotyping.

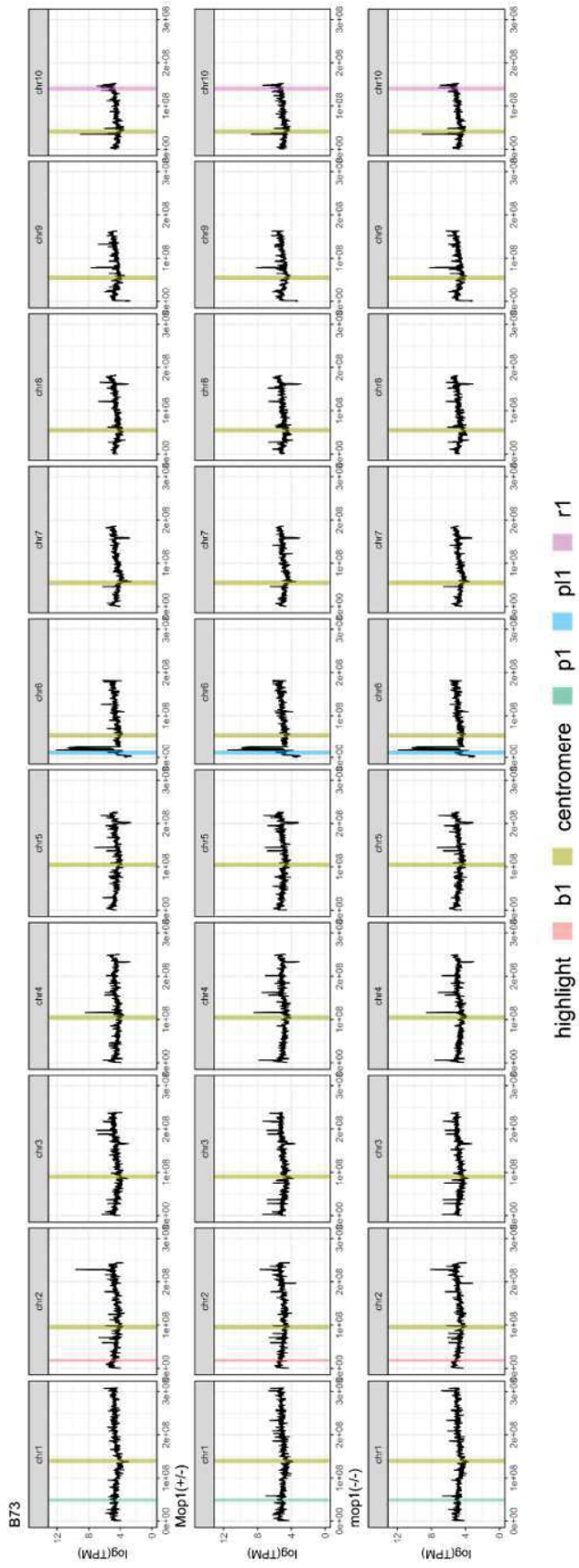



**Fig. S1** siRNA chromosome coverage on the B73 reference genome (version 5). siRNAs were extracted from AGO104 IPs in immature ears of three genetic backgrounds (B73, *Mm* and *mm*) with 2 technical and biological repeats. Reads were normalized in each sample using the TPM procedure. Colored highlights are the positions of the centromeres and the four known paramutation loci in maize (*p1* on chromosome 1, *b1* on chromosome 2, *pl1* on chromosome 6 and *r1* on chromosome 10).

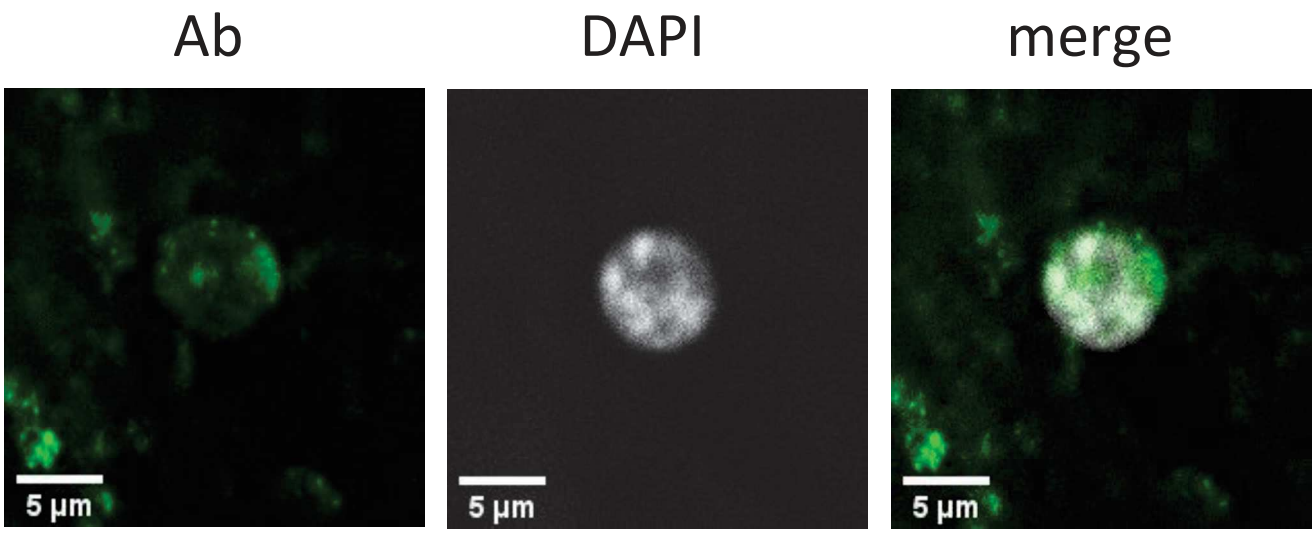



**Fig. S2** Fluorescence of AGO104 in nuclei of B73 embryonic cells 3 days after pollination (DAP). N=2 embryos and >100 cells show the same profile.

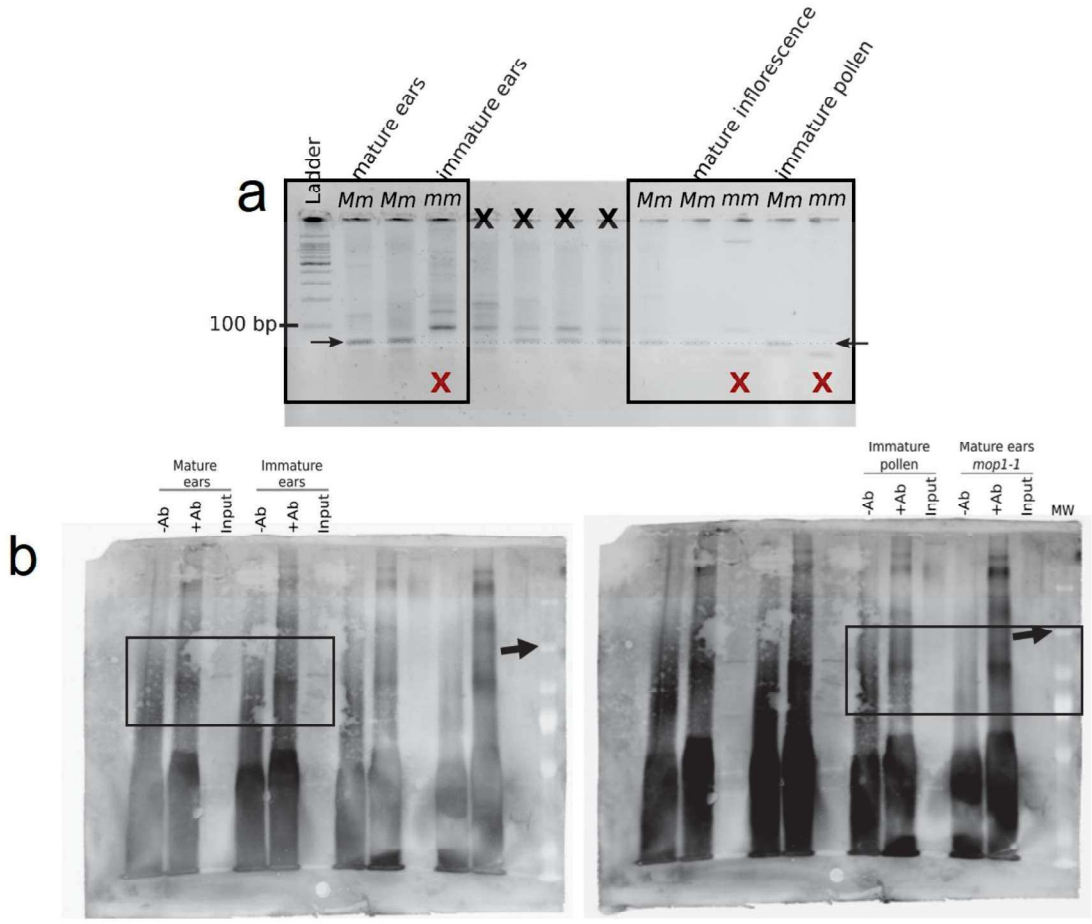



**Fig. S3** Complete gels presented in Fig. 1. (a) Stem loop gel presented in Fig 1b. Rectangle indicates the region presented. Arrow stand for the expected bands at 67 bp. Red cross is indicated when no R3 siRNA is detected. (b) Western blot presented in Fig 1b. Rectangle indicates the region presented. Arrow stands for 100 kD.

**Table S1** AGO104 antibody characteristics.

| Target protein | Target peptide | Peptide position | Organism |
|---|---|---|---|
| AGO104 | SERICKEQTFPLRQR | 326-341 | Rabbit |

**Table S2** Primer sequences used for R3 siRNAs stem loop PCR.

| siRNA sequence | Stem loop primer | TM | Forward primer | TM |
|---|---|---|---|---|
| ATGATTTGTGGGTCCGATGGCATA | GTCGTATCCAGTGCAGGGTCCGAG GTATTCGCACTGGATACGATATGCC | 73 | TGCCGATGATTTGTGGGTCC | 59 |

**Table S3** Primer sequences used for genotyping. The *Mutator* primer was associated with the Forward primers.

| | Forward | Reverse | Mutator |
|---|---|---|---|
| *mop1-1* | TCTCCACCGCCCACTTGAT | ATGGCCAGCAGGGT GTCGCAGAT | AGAGAAGCCAACGCCA WCGCCTCYATTTCGTC |
| *ago104-5* | TGTCTCCTGTATCAACGGGGTGGTC | CTATACCAGGCCTGT CAATCAGTAATCTC | |

# Chapter 2

# Identification of new paramutation loci in *Zea mays*

In the second chapter of my PhD thesis, I develop a method to explore the extent of the paramutation phenomenon in maize. Indeed, determining whether paramutation is a global silencing mechanism remains an intriguing, unanswered question. To address this, I selected distant inbred lines and generated reciprocal F1s and several backcrossing generations. At each generation, I performed genome wide mRNA sequencing and I designed a pipeline to search for gene expression patterns matching the three steps of paramutation described by Brink, Coe and Hagemann (see General Introduction). In total, I identified 147 candidate genes that fit these three key steps. I therefore demonstrated that it is possible to identify new loci subjected to paramutation, and that paramutation is more common than initially anticipated in maize. Then, I used these new loci to search for common features that could help towards understanding both the origin and the molecular nature of the mechanisms mediating paramutation. These genes showed no trend for function and typical genomic features. In addition, none but four of them unexpectedly did match the genes which expression id dependent on RdDM members (ie, MOP1 and MOP3). Therefore, further research will be needed to identify the factors responsible for triggering paramutagenicity in maize.

**My contribution**

I produced all the research presented in this chapter (including all the crosses presented, RNA extraction and RNA Illumina libraries), and analysed the data by producing all the scripts required. I presented this work at the 2019 European Maize Meeting (Montpellier, France), at the 2021 Maize Genetics Meeting (63$^{rd}$ Annual Maize Genetics Meeting, Virtual), and as an invited speaker at the Annual INDEPTH Meeting 2021.

This chapter is presented as a manuscript and will be submitted for publication in forthcoming weeks.

# Chapter 2 Index

# Identification of new paramutation loci in Zea mays

Juliette Aubert [1], Mario A. Arteaga-Vazquez [2], Olivier Leblanc [1], Daniel Grimanelli [1]

[1] DIADE, University of Montpellier, CIRAD, IRD, Montpellier, France

[2] Universidad Veracruzana, INBIOTECA, Xalapa, Veracruz, Mexico

## Abstract

Paramutation is a rare exception where reprogramming is both mitotically and meiotically stable over many generations. Since its discovery in 1956, only 4 loci subjected to paramutation have been described in *Zea mays*, all associated with a red pigmentation phenotype: *red1* (*r1*), *booster1* (*b1*), *plant color1* (*pl1*), and *pericarp color1* (*p1*). To determine the extent of paramutation in gene silencing, we selected three genetically distant inbred lines (B73, M37W and M162W) to generate a crossing scheme aimed at fostering the detection of loci subjected to paramutation using gene expression as a phenotypic readout. We sequenced and analysed leaf mRNA in four successive generations, and identified 147 differentially expressed genes in the inbred lines that were stably silenced in both F1s and backcross generations. These alleles met all criteria required for paramutation as silencing is stable through meiosis and the newly silenced alleles are capable of paramutagenic activity in subsequent generations. They cover diverse functions in maize, but lack exhibiting both genomic and functional similarities, including those typically observed for known loci. Based on these results, we argue that the paramutation phenomenon is more common than initially expected in maize and it should not be overlooked when crossing independent lines.

## Introduction

Progress in RNA sequencing has been a major goal since it proved useful in improving agronomic traits. Analyses were conducted to inventory gene expression patterns in the progeny of plants with differentially expressed genes (DEG) using RNA seq (Stupar and Springer, 2006; Li et al., 2013). In maize, 80% of parental DEGs show an additive expression pattern in F1 hybrids, meaning that 20% of those parental DEG behave differently than expected in the hybrids and could not be predicted by Mendelian inheritance laws (Stupar and Springer, 2006). These 20% are often within parental range of expression in the hybrids, but some display extreme expression patterns, either above the highly expressed parent or below the lowly expressed parent. Among these unexpected patterns, some transcripts are completely absent in the hybrids. Other interesting patterns were identified, such as paramutation-like behaviour with genes being expressed similarly to the low parent in all hybrids (Li et al., 2013).

Paramutation is an intriguing epigenetic phenomenon first observed in the 1910's (for an historical perspective, see see (Chandler and Stam, 2004)) and later defined by Brink (Brink, 1956) at the *red1* (*r1*) locus in maize. Next, the phenomenon was detected in several species (Brink, 1956; Rassoulzadegan et al., 2006; De Vanssay et al., 2012; Sapetschnig et al., 2015) and at three other maize loci, including *booster1* (*b1*) (Coe, 1959), *plant color1* (*pl1*) (Hollick et al., 1995), and *pericarp color1* (*p1*) (Sidorenko and Peterson, 2001). Paramutation involves the establishment of heritable changes in gene expression following trans-silencing interactions triggered by a silenced, paramutagenic allele and targeting a paramutable allele. In addition, newly silenced alleles also acquire paramutagenic capabilities when combined with a paramutable allele resulting in secondary paramutation. These three steps define paramutation as described when discovered (Brink, 1956; Coe, 1959; Hagemann, 1969). The level of expression of the four maize loci known to be subjected to paramutation enables to determine their paramutation state using anthocyanin production as a read out (Chandler et al., 2000). While this behaviour is dependent on the RdDM pathway for the four loci (Dorweiler et al., 2000), diverging features have been reported. At *b1*, the *B'* (paramutagenic) allele was never observed to revert to its *B-I* (paramutable) state even after several backcrosses to *B-I*, whereas high rates of reversion are observed for paramutagenic alleles at *r1* and *pl1*, respectively *R-rs* and *Pl'*. Furthermore, *B-I* shows spontaneous switch to its silenced *B'* state more often than the other loci (reviewed in (Chandler et al., 2000)), and shows no intermediate pigmentation as compared to the *r1* or *p1* loci (Coe, 1966; Sidorenko and Peterson, 2001). Finally, alleles engaged in paramutation exhibit genomic rearrangements possibly involved in trans-silencing interactions through the RNA-directed DNA methylation (RdDM) pathway, including repeated sequences, transposable elements (TEs) and tandem

duplications (Kermicle et al., 1995; Sidorenko and Peterson, 2001; Stam et al., 2002b; Giacopelli and Hollick, 2015).

Examples of paramutation or paramutation-like behaviours have been reported in several other eukaryotes, as well as a pivotal role for small RNA effectors, TEs and regulators of genome stability (Gabriel and Hollick, 2015). However, there are mechanistic differences between plants and animals (Chandler, 2007; Gabriel and Hollick, 2015) and, in maize, the resulting alterations in gene expression vary greatly depending on the locus, ie from on/off states to more subtle variations. The extent of the phenomenon in metazoans remains unknown as well as its origin and functional role. In maize, Li et al. (Li et al., 2013) reported 145 genes showing expression patterns reminiscent of paramutation in a study RIL genotypes generated using the B73 and Mo17 inbreds. Determining whether paramutation acts sporadically or extensively affects gene expression is a key to define the underlying mechanisms and to understand its role in establishing gene expression patterns during evolution.

To answer that question in maize, we used RNAseq to characterize gene expression in F1 and backcross derivatives obtained with the B73 inbred and two distant South African inbreds, M37W and M162W. Using a RNAseq pipeline for processing read alignments and monitoring gene expression patterns, we identified loci possibly subjected to paramutation and searched for common genomic and genic features. Although we found 147 genes, our results suggest that paramutation strongly depends on genotype interactions rather than common features between these 147 genes.

## Results

### Identification of new genes subjected to paramutation

Paramutated alleles are detectable when they are differentially expressed in parents. One allele is silenced by its counterpart in the F1, meaning that paramutated genes are expressed below the expected mid-parent expression in the F1. After several backcrosses, paramutated gene expression may vary between the four known loci of paramutation in maize. We designed our study to reveal paramutation events following the *b1* model, which is particularly convenient to monitor gene expression levels across generations (F1, BC1, BC2) as it has the most extreme silencing behaviour compared to the other three loci. *B'/B-I* and *B'/B'* have similar pigmentation resulting from similar expression of the paramutagenic allele and the recently converted allele (Patterson et al., 1993). The *b1* paramutagenic state is similar to the lowly expressed parent and is quantifiable using gene expression. The *B'* expression level is highly stable and can be monitored across generations, meaning that expression of the low parent is equal to the expression in the F1s, and successive BC generations.

We designed a procedure to identify similar gene expression using parental DEG that are expressed below the expected mid-parent expression in the F1, and maintain their weak expression after backcrosses (**Fig. 1A-B**).

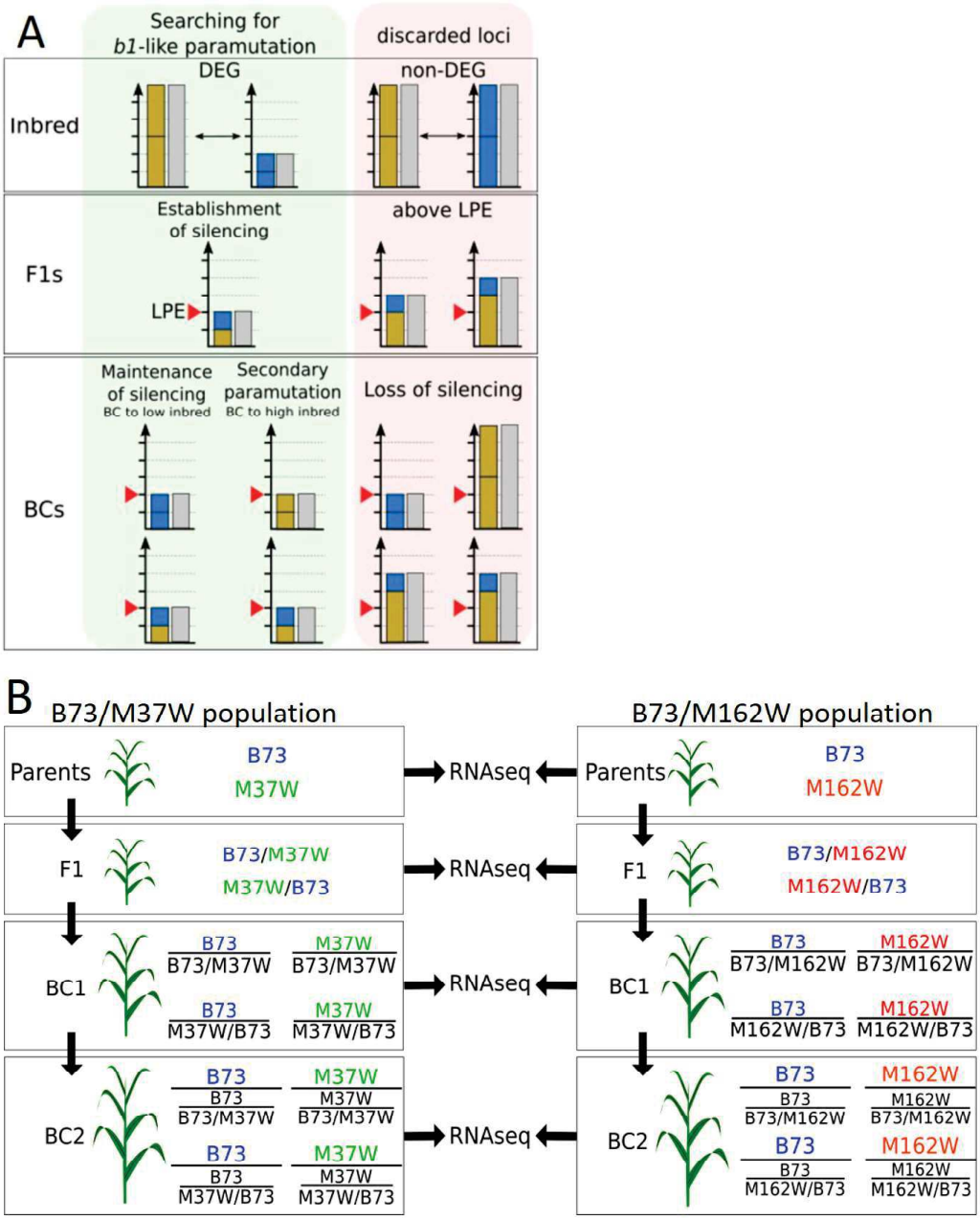



**Figure 1.** Creation of populations to select candidates to paramutation. A) Selection of candidates to paramutation using RNAseq data. Height of grey rectangles is the level of expression of the selected gene. Blue and brown rectangles are parental alleles, and their height represent their involvement in gene expression. LPE is the level of expression of the lowly expressed parent (Low Parent Expression). B) Crossing scheme from inbred parents to BC2. The left population is derived from B73 with M37W, and the right population is derived from B73 with M162W. Messenger RNA from the 4th leaf was sequenced for each genotype, with 3 technical and biological repeats for each.

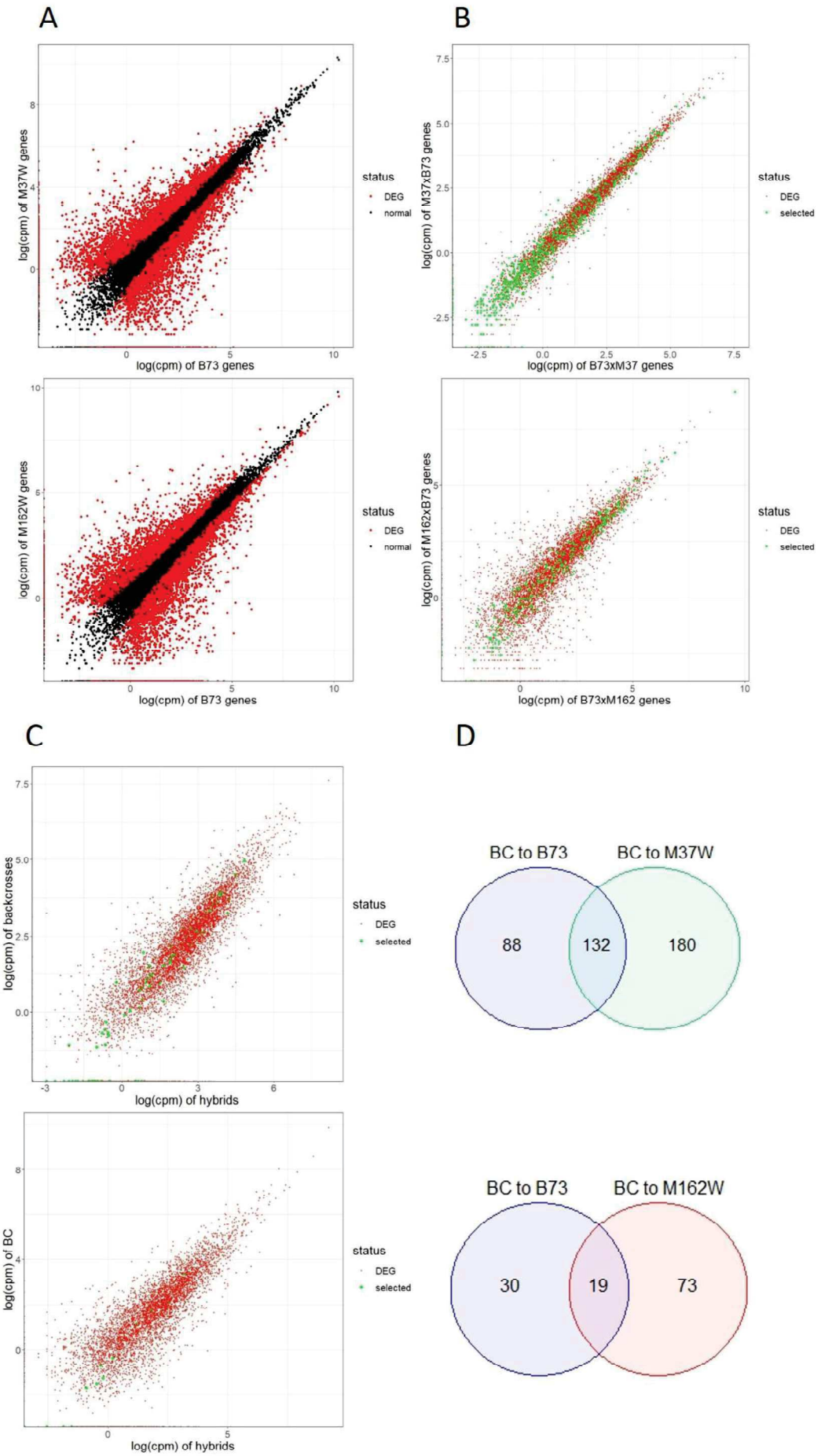

A
log(cpm) of M37W genes
log(cpm) of B73 genes
status
DEG
normal
B
log(cpm) of M37xB73 genes
log(cpm) of B73xM37 genes
status
DEG
selected
log(cpm) of M162W genes
log(cpm) of B73 genes
status
DEG
normal
log(cpm) of M162xB73 genes
log(cpm) of B73xM162 genes
status
DEG
selected
C
log(cpm) of backcrosses
log(cpm) of hybrids
status
DEG
selected
D
BC to B73
BC to M37W
88
132
180
log(cpm) of BC
log(cpm) of hybrids
status
DEG
selected
BC to B73
BC to M162W
30
19
73

**Figure 2.** Gene expression in inbred lines and in subsequent generations. A) DEG in the parental inbred lines. Black dots are all B73 genes, and red dots are DEG. B) Selection of the parental DEG that have a similar expression to the low expressed parent in both reciprocal crosses. Red dots are parental DEG and green dots are candidate genes, selected after the F1 crosses. C) Selection of parental DEG that continue to have the same expression as the weakly expressed parent in all backcrosses. Red dots are parental DEG and green dots are candidate genes, selected after the BC2 backcrosses. D) Selected genes in hybrid backcrossed to B73 (blue circle), to M37W (green circle), or to M162W (red circle).

Using RNA sequencing (RNAseq), we monitored gene expression throughout generations in individuals derived from crosses between maize inbred lines. To select for parental inbreds, we rationalized that maximizing the divergence between the parental inbreds would favor the discovery of alleles involved in paramutation. Maximizing allelic variation enables to increase the number of parental DEGs, which is a prerequisite for identifying paramutation behavior. We selected two South-African inbred lines, M37W and M162W, that were previously described as distant from the B73 inbred line (Yu et al., 2008). We then sequenced RNA from the 4$^{th}$ leaf of all three inbred lines and identified 7 129 and 6 123 DEGs, respectively between B73 and M37W, and B73 and M162W (**Fig. 2A**).

We then searched for parental DEG that are expressed below mid-parent value in F1s derived from reciprocal crosses between B73 and M37W, and B73 and M162W. We sequenced RNA from the 4$^{th}$ leaf of F1 plants and selected parental DEGs that showed similar expression levels to that of the low expressed parental allele. This expression pattern results from the silencing of highly expressed alleles. Among the selected candidates, two groups can be designed: alleles that were highly expressed in the South African inbred line and silenced by the B73 alleles in hybrids (group A), and alleles that were highly expressed in B73 and silenced by the South African inbred alleles in hybrids (group B). In total, 1 486 (824 in group A and 662 in group B) and 975 (826 in group A and 149 in group B) genes were retained in F1 progeny derived from M37W and from M162W, respectively (**Fig. 2B**). This behavior, similar to that of *b1,* is indicative of the establishment of paramutation when alleles subjected to paramutation are combined. However, at this point, candidate genes cannot be defined as "paramutable" as their ability to maintain and establish secondary paramutation was not assessed yet.

Using *b1* as a model, maintenance of paramutation is the ability of newly silenced *B'* alleles to remain silenced when the original paramutagenic allele is segregated away. We therefore evaluated expression level of the selected candidate genes throughout two successive backcross generations (BC1 and BC2) derived from each reciprocal F1. We performed each backcross three times which

yielded three cobs, from which we pooled and sequenced the RNA of three plants each (three libraries and nine plants sequenced per backcross). For all subsequent analysis, it is very unlikely that any dominant repressor was not segregated away at least in one library, and we therefore excluded the hypothesis of a dominant repressor controlling gene expression. At each generation, we used RNAseq from the 4$^{th}$ leaf to filter loci that remained silenced or showed expression levels in BC1 and BC2 plants similar to that of the low parental allele. We verified that group A candidates (B73 carries the lowly expressed allele) maintained their weak expression in all BC1 and BC2 that were backcrossed to B73. Similarly, we verified that candidates from group B (South African inbred lines carry the lowly expressed allele) maintained their weak expression in all BC1 and BC2 that were backcrossed to the South African inbred lines. This allowed the selection in total of 220 genes from the B73/M37W population and 49 genes from the B73/ M162W population. Those genes are capable to maintain silencing through generations and, therefore, fit the definition of maintenance of paramutation (**Fig. 2C**).

Furthermore, newly silenced *B'* alleles also acquire paramutagenic activity, called secondary paramutation, the key feature for defining paramutation. This means that newly silenced *B'* alleles produce a *B'* progeny when backcrossed to the highly expressed *B-I* allele. We therefore used the RNAseq generated from the previous backcrosses derived from each reciprocal F1. We verified that candidates from group A (B73 carries the lowly expressed allele) were still silenced in all BC1 and BC2 that were backcrossed to the South African inbred lines (carrying the highly expressed allele). We conducted the same analysis by verifying that candidates from group B (South African inbred lines carry the lowly expressed allele) were still silenced in all BC1 and BC2 that were backcrossed to the B73 inbred (carrying the highly expressed allele). All candidates that validated these steps are therefore capable of secondary paramutation. Out of the 220 genes capable of maintaining silencing in the B73/M37W population, 132 are also capable of secondary paramutation (**Fig. 2D**). In the B73/ M162W population, 19 genes out of the 49 previously identified are capable of secondary paramutation (**Fig 2D**). As four genes were found in common between the M37W and M162W populations, we identified a total of 147 genes that do secondary paramutation. These 147 genes behave identically to *b1* in a paramutagenic context as they can establish and maintain paramutation, as well as induce secondary paramutation.

## Candidate genes intersect with known gene sets

These candidate genes provide a new perspective to identify shared patterns among paramutable alleles and, hopefully, help in understanding what causes paramutation. We firstly evaluated gene

ontology for all the candidate genes and identified no gene enrichment for specific functions. In maize, MEDIATOR OF PARAMUTATION1 (MOP1) and MOP3, two components of the RdDM pathway, are involved in paramutation at the four known loci, and also control the expression level of many other genes in B73 (Madzima et al., 2014; Forestan et al., 2017). We compared those Mop1 and Mop3-dependant genes to the set of candidate loci we identified with paramutagenic activity from the B73 allele (ie South-African alleles that are silenced by the B73 alleles). Under our hypothesis, we expect that both *mop1* and *mop3* mutations release the silencing of the paramutagenic B73 alleles. Interestingly, out of the 61 candidates from the cross of B73 with M37W and four from B73 with M162W, only 1 locus (Zm00001d017700/*JUMONJI-transcription factor 15*) matched the *mop1*-dependent list and none matched the *mop3*-dependent list. Overall, comparisons of candidate genes obtained from BC2 plants backcrossed twice to B73 and twice to the South-African parental lines with the *mop1*-dependent and *mop3* -dependent genes identified few genes. Therefore, such weak overlapping between the lists of genes is poorly explained by the difference of genetic background. We obtained similar results using upregulated genes in *zmet2 and zmet5* chromomethylase mutants (genes that escape silencing by demethylation) (Anderson et al., 2018), as we found only one candidate reported for the *zmet2-2* mutant (Zm00001d034357). These results seem to indicate that paramutation candidate genes are not dependant on previously identified actors of paramutation. Finally, none of the candidates we identified matched the 145 genes identified in B73xMO17 RILs (Li et al., 2013). Interestingly, those 145 genes do not match the list of MOP1-dependant genes either, which strengthens our results.

## Genomic features of the candidate genes

We hypothesized that paramutation-like behaviour in nearby genes might be acquired through spreading of the establishment of paramutation at one locus. We therefore looked for candidate genes distribution along the 10 maize chromosomes by testing the proportion of candidate genes to the total number of expressed genes in 5 Mb windows against the same ratio along the chromosome. We noted that the candidate genes did not cluster around the previously known loci of paramutation and were scattered on all 10 chromosome arms, reminding the position of the B73 genes along chromosomes. We created 5 Mb windows along the B73 genome and compared the number of expressed genes in B73 with the number of candidate genes in each window. All proportion tests indicated no statistical difference in repartition between windows ($p > .05$) nor between chromosomes ($p = .28$). This result implies that genes regulated by paramutation can be located on any of the 10 chromosomes.

Paramutation requires the presence of seven TRs 100 kb upstream of the *b1* locus (Stam et al., 2002b) and the presence of TRs and TEs upstream of the *p1* locus (Wang et al., 2017), independently of their paramutation state. Therefore, we searched for loci with higher density of repeat sequences or TEs up to 100 kb upstream from the terminator of each candidate gene. To determine whether density was higher than average, we selected 100 genes randomly and ran the same analysis. We found no significant difference between the control and the candidate genes groups. We narrowed the window to the 20 kb upstream and down to the terminator of both candidate genes and control genes as an attempt to increase the density of regulatory elements per genes (Ricci et al., 2019). Again, we found no significant difference for repeats or TEs density (**Fig. 3A-B**), suggesting that alleles with a paramutation-like behaviour are not regulated by different densities of TEs or repeats any more than other genes.

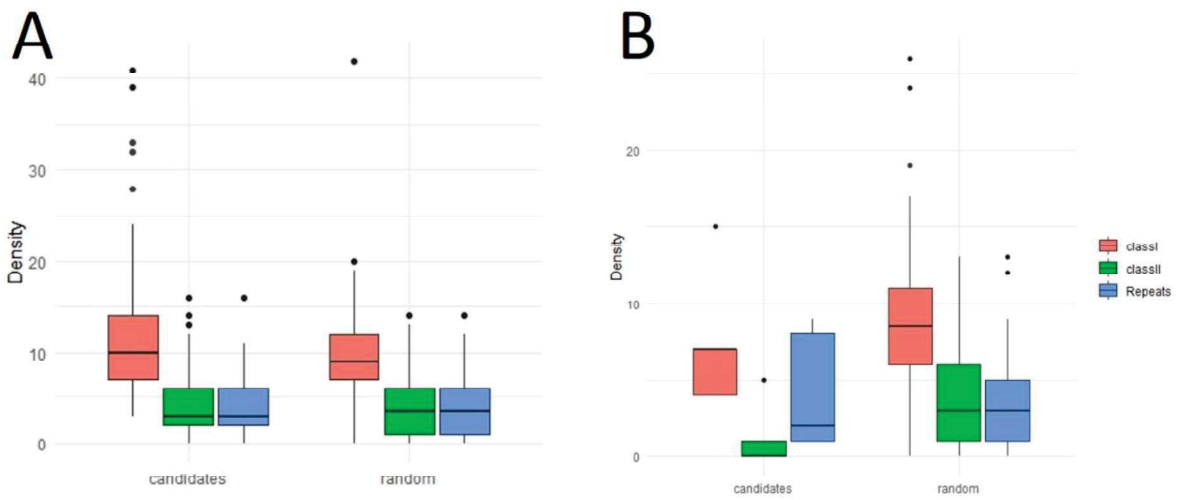



**Figure 3.** Comparison of TE density between random genes and candidate genes (selected at BC2) **A)** in the M37W population and **B)** the M162W population. Density is shown for region spanning 20kb upstream the terminator of the genes. TEs and repeats are separated by class.

## Discussion

Paramutation is not the only example of phenomenon that alter gene expression through generations without changes in the genetic code. Imprinting has similarities with paramutation as it silences genes in hybrids to reach the expression level of the low expressed parent (Kermicle and Alleman, 1990). It is thought to occur only in endosperm of maize, but it is not possible to exclude its involvement in the regulation of the candidate genes in the F1. However, we performed reciprocal crosses and backcrosses which enables to exclude the involvement of imprinting in the regulation of the candidate genes.

In the *B'* allele, the expression of the *b1* gene is lower than in the *B-I* allele, but it is not completely switched off (Patterson et al., 1993). Therefore, the filter we used for gene selection conserves candidate genes that are switched off or simply partially silenced. Our strategy allowed the identification of 147 genes, for which expression patterns are in agreement with paramutation at the *b1* locus. Although this represent a small fraction of expressed genes, it is worth noting that our screen was based on restrictive criteria, following the most extreme paramutation model identified in maize. Thus, it is likely that paramutation events are wider than those adjusting to the *b1* model, an assumption supported by Li et al. (2013) results showing that about half (55%) of the candidate genes for paramutation had expression levels similar to the lower expression parent. Furthermore, the lack of overlap between the three different sets of genes currently available for paramutation (this work and (Li et al., 2013)) all involving B73 alleles, is unexpected. This may indicate either that our criteria were too stringent to select paramutable genes, or that paramutable alleles vary drastically between inbred lines.

In addition to determine the extent of the paramutation phenomenon, we also searched for functional and genomic features that could shed light on underlying regulatory mechanisms. First, in contrast with the four known loci in maize involved in anthocyanin biosynthesis, the candidate genes did not belong to the list of genes regulated by either MOP1 or MOP3 (Madzima et al., 2014; Forestan et al., 2017) nor by the zmet2 and zmet5 methyltransferases. This suggests that the RdDM pathway does not regulate the candidate genes until they are crossed with plants carrying the corresponding paramutagenic allele. Candidate genes seem to have different properties from the four known loci of paramutation, although it is possible that our stringent criteria for selection of candidates eliminated potential candidates that are regulated by RdDM.

We next focused our analysis on locating and identifying regulatory regions as observed for the four known loci subjected to paramutation. These are expected to be found in enhancer regions, located upstream of the genes and containing TRs or TEs (Stam et al., 2002a; Stam et al., 2002b; Wang et al., 2017), and involved in the synthesis of 24-nt siRNAs through the RdDM pathway (Arteaga-Vazquez and Chandler, 2010) although the presence of TRs or TEs around candidate genes does not imply that they consist in enhancers of paramutation, as 85% of the maize genome is made of these repeated sequences (Jiao et al., 2017). Our statistical analysis showed no evidence to support a difference for repeats and TEs density around candidate genes.

Overall, our results show that paramutation is more common than initially expected in maize, but is probably not a global silencing mechanism as only 132 genes in the M37W crosses and 19 genes in the M162W crosses were identified. These candidate genes were obtained after sequencing RNA from four successive generations with reciprocal backcrosses and applying reliable statistical filters to monitor gene expression. The lack of common feature and of evidence for the role of RdDM in candidate genes regulation calls for further research to understand the mechanisms responsible for establishing paramutagenic activity.

## Materials & methods

### Plant stock & growth conditions

B73, M162W and M37W were provided by the Maize Genetics Cooperation Stock Center. Inbred and derived individuals were grown in a greenhouse at the French National Research Institute for Sustainable Development in Montpellier, France, with 14 hours daylight at 26°C and 60% humidity. The 4$^{th}$ leaf was collected in triplicate and immediately snap frozen in liquid nitrogen and stored at -80°C before use. Triplicates were obtained from seedlings from different cobs, but each replicate pooled three seedlings from the same cob.

### Obtaining mRNA libraries

Total RNA was extracted using Tri Reagent solution following Applied Biosystems recommendations. mRNA extraction from total RNA was performed using the Thermo Scientific MagJET mRNA Enrichment Kit following manufacturer instructions for manual mRNA purification, except for the final elution that was done using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina. As advised in the instructions manual (at step 1.2.36.), we added 11.5 µl of the First Strand Synthesis Reaction Buffer and Random Primer Mix. PCR enrichment was performed using 6 to 9 cycles depending on the initial amount of total RNA. Samples were quantified using Qubit and a Bioanalyzer Agilent High Sensitivity DNA Chip and sequenced on a NextSeq550 machine at the CSHL Genome Center.

### Data processing

We sequenced 69 libraries, that consisted in 3 technical and biological repeats of 23 libraries. There is 12 libraries for the crosses of B73 with M37W and 12 libraries for the crosses of B73 with M162W (**Fig. 1A**). The B73 genotype is present in both populations but was only sequenced once, hence 23 libraries were sequenced. Raw sequencing data were cleaned using Trimmomatic (version 0.38) with parameters LEADING:10 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Reads were aligned to the

B73 reference genome version 5 using HISAT2 (version 2.1.0) with the following parameters –k 2 -3 3 -5 8, and reads with multiple alignments were discarded. We identified gene IDs by checking overlapping of reads with the transcripts from the B73 reference gff file version 5 using Bedtools intersect (version 2.26.0 (Quinlan and Hall, 2010)). We used EdgeR (Robinson et al., 2009) for data normalization and discarded lowly expressed genes (averaged expression < 1 count per million, in both parents). The exactTest function from the EdgeR package was used to determine DEG between the parental lines and to select candidate genes that are differentially expressed from the highly expressed parent but non-differentially expressed from the lowly expressed parent.

### Bioinformatics analysis

### Mutant dependant genes

The list of genes showing an expression pattern dependant on Mop1, Zmet2 and Zmet5 were all created in the B73 genetic background (Madzima et al., 2014; Anderson et al., 2018) whereas that for Mop3 was obtained from an undescribed genetic background (Forestan et al., 2017). When needed, gene IDs (Version5) were converted into Version3 or Version4 IDs using the "Translate Gene Model IDs" tool available at maizegdb.org.

### Chromosomal distribution of candidate genes and nearby TE density

We used the prop.test function of R for testing candidate genes distribution among chromosomes and the proportion of candidate genes vs. expressed genes in 5Mb windows against the same proportion observed at the chromosome level. All *p* values were adjusted using mt.rawp2adjp multiple testing function of R (Benjamini and Hockberg procedure). Windows and counts per window were obtained using the Bedtools makewindows and intersect commands (Quinlan and Hall, 2010). Global TE and repeat annotations were downloaded from the output of the RepeatMasker's analysis on B73 reference genome version 5 on NCBI. Densities of TEs and repeats on the randomly selected genes and on the candidate genes was performed using the GenomicRanges package (Lawrence et al., 2013) that identifies overlaps between lists of genomic positions.

### Data availability

All RNAseq data will be made available on GEO before publication.

## Fundings

D.G. received support from the H2020 MCA (REP-658900-2) and *Agence Nationale de la Recherche* grants REMETH (ANR-15-CE12-0012-03). D.G. and O.L. received support from the CHROMOBREED

(ANR-18-CE92-0041). M.A.A.V. received support from the *Jeunes équipes associées à l'IRD (JEAI)* program (EPIMAIZE), Agropolis Foundation, CONACYT (158550 & A1-S-38383) and the Royal Society Newton Advanced Fellowship (NA150181).

## Author contributions

J.A., O.L., M.A.A.V., D.G. designed the research; J.A, D.G. performed the research; J.A., O.L., D.G. analysed the data; D.G. acquired funding; all authors wrote the paper.

**Competing Interest Statement:** The authors declare no competing interest.

## References

**Anderson SN, Zynda GJ, Song J, Han Z, Vaughn MW, Li Q, Springer NM** (2018) Subtle perturbations of the maize methylome reveal genes and transposons silenced by chromomethylase or RNA-directed DNA methylation pathways. G3 Genes, Genomes, Genet **8**: 1921–1932

**Arteaga-Vazquez MA, Chandler VL** (2010) Paramutation in maize: RNA mediated trans-generational gene silencing. Curr Opin Genet Dev **20**: 156–163

**Brink RA** (1956) A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. Genetics **41**: 872–89

**Chandler VL, Eggleston WB, Dorweiler JE** (2000) Paramutation in maize. Plant Mol Biol **43**: 121–145

**Chandler VL, Stam M** (2004) Chromatin conversations: Mechanisms and implications of paramutation. Nat Rev Genet **5**: 532–544

**Coe EH** (1966) THE PROPERTIES, ORIGIN, AND MECHANISM OF CONVERSION-TYPE INHERITANCE AT THE *B* LOCUS IN MAIZE. Genetics **53**: 1035–1063

**Coe EH** (1959) A Regular and Continuing Converstion-Type Phenomenon at the B Locus in Maize. Proc Natl Acad Sci 828–832

**Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL** (2000) *mediator of paramutation1* Is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci. Plant Cell **12**: 2101–2118

**Forestan C, Farinati S, Aiese Cigliano R, Lunardon A, Sanseverino W, Varotto S** (2017) Maize RNA PolIV affects the expression of genes with nearby TE insertions and has a genome-wide repressive impact on transcription. BMC Plant Biol **17**: 1–27

**Gabriel JM, Hollick JB** (2015) Paramutation in maize and related behaviors in metazoans. Semin Cell Dev Biol **44**: 11–21

**Giacopelli BJ, Hollick JB** (2015) Trans-Homolog interactions facilitating paramutation in maize. Plant Physiol **168**: 1226–1236

**Hagemann R** (1969) SOMATIC CONVERSION (PARAMUTATION) AT THE SULFUREA LOCUS OF LYCOPERSICON ESCULENTUM MILL. III. STUDIES WITH TRISOMICS . Can J Genet Cytol **11**: 346–358

**Hollick JB, Patterson GI, Coe EH, Cone KC, Chandler VL** (1995) Allelic interactions heritably alter the activity of a metastable maize pl allele. Genetics **141**: 709–719

**Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, et al** (2017) Improved maize reference genome with single-molecule technologies. Nature **546**: 524–527

**Kermicle JL, Alleman M** (1990) Gametic imprinting in maize in relation to the angiosperm life cycle. Dev Suppl 9–14

**Kermicle JL, Eggleston WB, Alleman M** (1995) Organization of paramutagenicity in R-stippled maize. Genetics 141(1) 361-372

**Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ** (2013) Software for Computing and Annotating Genomic Ranges. PLoS Comput Biol **9**: e1003118

**Li L, Petsch K, Shimizu R, Liu S, Xu WW, Ying K, Yu J, Scanlon MJ, Schnable PS, Timmermans MCP, et al** (2013) Mendelian and Non-Mendelian Regulation of Gene Expression in Maize. PLoS Genet **9**: e1003202

**Madzima TF, Huang J, McGinnis KM** (2014) Chromatin structure and gene expression changes associated with loss of MOP1 activity in Zea Mays. Epigenetics **9**: 1047–1059

**Patterson GI, Thorpe CJ, Chandler VL** (1993) Paramutation, an allelic interaction, is associated with a stable and heritable reduction of transcription of the maize b regulatory gene. Genetics 135:

**Quinlan AR, Hall IM** (2010) BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841–842

**Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F** (2006) RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. doi: 10.1038/nature04674

**Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-guerra MK, Colomé-tatché M, et al** (2019) Widespread long-range cis-regulatory elements in the maize genome. Nat Plants **5**: 1237–1249

**Robinson MD, McCarthy DJ, Smyth GK** (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**: 139–140

**Sapetschnig A, Sarkies P, Lehrbach NJ, Miska EA** (2015) Tertiary siRNAs Mediate Paramutation in C. elegans. PLoS Genet **11**: 1005078

**Sidorenko L V., Peterson T** (2001) Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize p1 gene. Plant Cell **13**: 319–335

**Stam M, Belele C, Dorweiler JE, Chandler VL** (2002a) Differential chromatin structure within a tandem

array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev **16**: 1906–1918

**Stam M, Belele C, Ramakrishna W, Dorweiler JE, Bennetzen JL, Chandler VL** (2002b) The regulatory regions required for B' paramutation and expression are located far upstream of the maize b1 transcribed sequences. Genetics 162(2) 917-930

**Stupar RM, Springer NM** (2006) Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics **173**: 2199–2210

**De Vanssay A, Bougé AL, Boivin A, Hermant C, Teysset L, Delmarre V, Antoniewski C, Ronsseray S** (2012) Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature **490**: 112–115

**Wang P-H, Wittmeyer KT, Lee T, Meyers BC, Chopra S** (2017) Overlapping RdDM and non-RdDM mechanisms work together to maintain somatic repression of a paramutagenic epiallele of maize pericarp color1. PLoS One **12**: e0187157

**Yu J, Holland JB, McMullen MD, Buckler ES** (2008) Genetic Design and Statistical Power of Nested Association Mapping in Maize. Genetics **178**: 539–551

## Supplemental data

Available supplemental data for this chapter:

**Supp. Table 1**. Illumina sequencing quality of crosses with the M37W inbred line.

**Supp. Table 2**. Illumina sequencing quality of crosses with the M162W inbred line.

**Supp. Table 3**. List of the 132 candidate genes selected at BC2 in the crosses with the M37W inbred line.

**Supp. Table 4**. List of the 19 candidate genes selected at BC2 in the crosses with the M162W inbred line.

**Supp. Table 1**. Illumina sequencing quality of crosses with the M37W inbred line.

| Generation | Cross | N reads | Aligned to B73 (%) |
|---|---|---|---|
| Parent | B73 | 41,386,856 | 88.42 |
| Parent | M37W | 26,852,898 | 89.53 |
| F1 | B73/M37W | 25,919,620 | 77.87 |
| F1 | M37W/B73 | 28,211,964 | 80.10 |
| BC2 | B73 / (B73/M37W) | 37,138,375 | 76.36 |
| BC2 | B73 / (M37W/B73) | 39,267,271 | 64.22 |
| BC2 | M37W / (B73/M37W) | 23,573,677 | 79.70 |
| BC2 | M37W / (M37W/B73) | 20,721,903 | 81.08 |
| BC3 | B73 / (B73 / (B73/M37W) ) | 37,195,319 | 74.01 |
| BC3 | B73 / (B73 / (M37W/B73) ) | 32,878,400 | 50.38 |
| BC3 | M37W / (M37W / (B73/M37W) ) | 32,579,518 | 79.46 |
| BC3 | M37W / (M37W / (M37W/B73) ) | 47,911,485 | 71.74 |

**Supp. Table 2**. Illumina sequencing quality of crosses with the M162W inbred line.

| Generation | Cross | N reads | Aligned to B73 (%) |
|---|---|---|---|
| Parent | B73 | 41,386,856 | 88.42 |
| Parent | M162W | 25,534,844 | 88.25 |
| F1 | B73/M162W | 22,140,441 | 79.21 |
| F1 | M162W/B73 | 32,068,565 | 80.47 |
| BC2 | B73 / (B73/M162W) | 36,834,674 | 72.68 |
| BC2 | B73 / (M162W/B73) | 45,844,102 | 60.99 |
| BC2 | M162W / (B73/M162W) | 35,279,134 | 79.48 |
| BC2 | M162W / (M162W/B73) | 36,241,492 | 72.56 |
| BC3 | B73 / (B73 / (B73/M162W) ) | 36,562,288 | 81.22 |
| BC3 | B73 / (B73 / (M162W/B73) ) | 39,845,844 | 62.47 |
| BC3 | M162W / (M162W / (B73/M162W) ) | 41,546,973 | 82.67 |
| BC3 | M162W / (M162W / (M162W/B73) ) | 38,424,183 | 82.10 |

**Supp. Table 3**. List of the 132 candidate genes selected at BC2 in the crosses with the M37W inbred line. Status column indicates which inbred parental allele downregulates the other one, and downregulates alleles in all subsequent generations.

| *genes* | chr | start | end | B73_mean_expr | M37_mean_expr | status |
|---|---|---|---|---|---|---|
| *Zm00001eb000820* | chr1 | 3022946 | 3028004 | 567 | 544.333333333333 | B73_down |
| *Zm00001eb016330* | chr1 | 57058410 | 57107779 | 1726.33333333333 | 3816.66666666667 | B73_down |
| *Zm00001eb020490* | chr1 | 76334885 | 76335744 | 9.33333333333333 | 61.6666666666667 | B73_down |
| *Zm00001eb032720* | chr1 | 182851043 | 182853377 | 29.3333333333333 | 96.6666666666667 | B73_down |
| *Zm00001eb042900* | chr1 | 226215534 | 226219226 | 188.333333333333 | 363.333333333333 | B73_down |
| *Zm00001eb050560* | chr1 | 257363226 | 257363860 | 23.3333333333333 | 87.6666666666667 | B73_down |
| *Zm00001eb050880* | chr1 | 259223818 | 259232407 | 544.333333333333 | 1078 | B73_down |
| *Zm00001eb059040* | chr1 | 289275840 | 289285660 | 120.333333333333 | 160.666666666667 | B73_down |
| *Zm00001eb059950* | chr1 | 292187637 | 292189587 | 141.666666666667 | 1260.66666666667 | B73_down |
| *Zm00001eb061440* | chr1 | 297022922 | 297031737 | 546.666666666667 | 609.666666666667 | B73_down |
| *Zm00001eb061470* | chr1 | 297134492 | 297142325 | 648 | 757 | B73_down |
| *Zm00001eb063210* | chr1 | 301896847 | 301904989 | 210.666666666667 | 385 | B73_down |
| *Zm00001eb066640* | chr2 | 2524918 | 2533246 | 311.666666666667 | 539.333333333333 | B73_down |
| *Zm00001eb066880* | chr2 | 2922762 | 2927275 | 7.33333333333333 | 69.6666666666667 | B73_down |
| *Zm00001eb074580* | chr2 | 20792127 | 20800612 | 86.3333333333333 | 170 | B73_down |
| *Zm00001eb077150* | chr2 | 29435103 | 29438161 | 8.66666666666667 | 31.6666666666667 | B73_down |
| *Zm00001eb078850* | chr2 | 35926173 | 35945022 | 995 | 1074 | B73_down |
| *Zm00001eb081600* | chr2 | 46786871 | 46789431 | 17 | 394.666666666667 | B73_down |
| *Zm00001eb085360* | chr2 | 72016320 | 72020650 | 329 | 785 | B73_down |
| *Zm00001eb089100* | chr2 | 114561277 | 114562516 | 8 | 81 | B73_down |
| *Zm00001eb093970* | chr2 | 156610575 | 156614700 | 1506 | 2185 | B73_down |
| *Zm00001eb094890* | chr2 | 162271068 | 162276069 | 466.333333333333 | 1638 | B73_down |
| *Zm00001eb106470* | chr2 | 210828889 | 210835136 | 13.6666666666667 | 96 | B73_down |
| *Zm00001eb110990* | chr2 | 222424609 | 222430276 | 128.333333333333 | 241 | B73_down |
| *Zm00001eb118220* | chr2 | 242703180 | 242719543 | 1851.33333333333 | 1875 | B73_down |
| *Zm00001eb119510* | chr3 | 2558624 | 2579876 | 1207.33333333333 | 3050.33333333333 | B73_down |
| *Zm00001eb125290* | chr3 | 22380002 | 22386372 | 108 | 132.333333333333 | B73_down |
| *Zm00001eb129120* | chr3 | 46466677 | 46470830 | 5 | 26.3333333333333 | B73_down |
| *Zm00001eb132590* | chr3 | 85695693 | 85696492 | 3.66666666666667 | 34 | B73_down |
| *Zm00001eb144330* | chr3 | 170735396 | 170757009 | 383 | 722.333333333333 | B73_down |
| *Zm00001eb146100* | chr3 | 177887899 | 177890153 | 0 | 34 | B73_down |
| *Zm00001eb148780* | chr3 | 187158633 | 187164353 | 361 | 443.666666666667 | B73_down |
| *Zm00001eb164200* | chr3 | 237537731 | 237538632 | 86 | 166.333333333333 | B73_down |
| *Zm00001eb179140* | chr4 | 82075313 | 82121098 | 1322.66666666667 | 2509.66666666667 | B73_down |
| *Zm00001eb180420* | chr4 | 94374535 | 94378852 | 330.666666666667 | 531.666666666667 | B73_down |
| *Zm00001eb188240* | chr4 | 164693645 | 164700513 | 300 | 516 | B73_down |
| *Zm00001eb213690* | chr5 | 5628700 | 5629887 | 0.666666666666667 | 106 | B73_down |
| *Zm00001eb257230* | chr5 | 220839482 | 220847359 | 59 | 248.333333333333 | B73_down |
| *Zm00001eb258080* | chr5 | 222544252 | 222549435 | 42 | 70 | B73_down |
| *Zm00001eb266920* | chr6 | 46438966 | 46441491 | 12 | 34.3333333333333 | B73_down |
| *Zm00001eb285190* | chr6 | 147363707 | 147368805 | 31.6666666666667 | 74 | B73_down |
| *Zm00001eb302550* | chr7 | 15376384 | 15377019 | 39.3333333333333 | 133 | B73_down |
| *Zm00001eb306430* | chr7 | 43404587 | 43435492 | 1590 | 1862.66666666667 | B73_down |
| *Zm00001eb315420* | chr7 | 133187931 | 133192951 | 1292.66666666667 | 2850.33333333333 | B73_down |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Zm00001eb320410* | chr7 | 151787132 | 151788726 | 32 | 173 | B73_down |
| *Zm00001eb334900* | chr8 | 11059736 | 11081360 | 650.666666666667 | 981.333333333333 | B73_down |
| *Zm00001eb351960* | chr8 | 121474574 | 121484366 | 3403.66666666667 | 5537.66666666667 | B73_down |
| *Zm00001eb356400* | chr8 | 139521543 | 139526556 | 7 | 32.3333333333333 | B73_down |
| *Zm00001eb359770* | chr8 | 153106153 | 153110291 | 48.6666666666667 | 119 | B73_down |
| *Zm00001eb366500* | chr8 | 172282801 | 172293334 | 511.666666666667 | 992.333333333333 | B73_down |
| *Zm00001eb367710* | chr8 | 174582146 | 174583745 | 11 | 70 | B73_down |
| *Zm00001eb367830* | chr8 | 174960987 | 174968833 | 196.666666666667 | 308.666666666667 | B73_down |
| *Zm00001eb380670* | chr9 | 39354777 | 39367826 | 323.333333333333 | 529.666666666667 | B73_down |
| *Zm00001eb395280* | chr9 | 137789171 | 137796963 | 144.666666666667 | 238.333333333333 | B73_down |
| *Zm00001eb396490* | chr9 | 142700397 | 142705748 | 130.666666666667 | 309.333333333333 | B73_down |
| *Zm00001eb414620* | chr10 | 71556508 | 71557504 | 1.66666666666667 | 26.3333333333333 | B73_down |
| *Zm00001eb414630* | chr10 | 71557839 | 71560434 | 6.33333333333333 | 87 | B73_down |
| *Zm00001eb419220* | chr10 | 98979398 | 98979819 | 0.666666666666667 | 32.6666666666667 | B73_down |
| *Zm00001eb423000* | chr10 | 121404151 | 121406464 | 1.33333333333333 | 25.6666666666667 | B73_down |
| *Zm00001eb429110* | chr10 | 140400313 | 140405095 | 218.333333333333 | 256 | B73_down |
| *Zm00001eb429450* | chr10 | 141418518 | 141421312 | 635 | 737 | B73_down |
| *Zm00001eb006120* | chr1 | 17513568 | 17516826 | 215.333333333333 | 14.6666666666667 | M37_down |
| *Zm00001eb031500* | chr1 | 176846939 | 176851816 | 60.6666666666667 | 4.66666666666667 | M37_down |
| *Zm00001eb032640* | chr1 | 182400738 | 182405039 | 2219.33333333333 | 679.333333333333 | M37_down |
| *Zm00001eb033880* | chr1 | 187771222 | 187771656 | 2426 | 88.6666666666667 | M37_down |
| *Zm00001eb041440* | chr1 | 218164563 | 218166155 | 193.333333333333 | 15 | M37_down |
| *Zm00001eb048410* | chr1 | 248672526 | 248673384 | 53.6666666666667 | 7 | M37_down |
| *Zm00001eb054930* | chr1 | 274379700 | 274382146 | 102 | 0 | M37_down |
| *Zm00001eb057000* | chr1 | 282273132 | 282275383 | 484.666666666667 | 120 | M37_down |
| *Zm00001eb059450* | chr1 | 290385257 | 290386856 | 166 | 0 | M37_down |
| *Zm00001eb062000* | chr1 | 298408180 | 298424189 | 159.666666666667 | 8.33333333333333 | M37_down |
| *Zm00001eb072080* | chr2 | 13537034 | 13538953 | 397.333333333333 | 61.6666666666667 | M37_down |
| *Zm00001eb082010* | chr2 | 48727359 | 48727970 | 43.6666666666667 | 0 | M37_down |
| *Zm00001eb082990* | chr2 | 54378800 | 54385584 | 43.3333333333333 | 0.333333333333333 | M37_down |
| *Zm00001eb087210* | chr2 | 94219971 | 94220460 | 101.666666666667 | 2.66666666666667 | M37_down |
| *Zm00001eb088790* | chr2 | 112273722 | 112279213 | 462.666666666667 | 3.66666666666667 | M37_down |
| *Zm00001eb089550* | chr2 | 116969095 | 116978844 | 134.333333333333 | 7 | M37_down |
| *Zm00001eb099110* | chr2 | 183824456 | 183837134 | 127 | 8 | M37_down |
| *Zm00001eb103600* | chr2 | 202366794 | 202371097 | 742 | 38.6666666666667 | M37_down |
| *Zm00001eb111030* | chr2 | 222456665 | 222460774 | 98.6666666666667 | 0.666666666666667 | M37_down |
| *Zm00001eb112140* | chr2 | 225616015 | 225621830 | 162.666666666667 | 1 | M37_down |
| *Zm00001eb112750* | chr2 | 227823952 | 227828785 | 12716 | 3925.33333333333 | M37_down |
| *Zm00001eb114370* | chr2 | 233095773 | 233101510 | 6235.33333333333 | 608.333333333333 | M37_down |
| *Zm00001eb116570* | chr2 | 238814044 | 238819930 | 1530.33333333333 | 133.666666666667 | M37_down |
| *Zm00001eb122750* | chr3 | 11618198 | 11625220 | 3931.66666666667 | 357.666666666667 | M37_down |
| *Zm00001eb123790* | chr3 | 15408971 | 15419667 | 1399.66666666667 | 516.333333333333 | M37_down |
| *Zm00001eb127420* | chr3 | 34673222 | 34674967 | 555.333333333333 | 122 | M37_down |
| *Zm00001eb137470* | chr3 | 133684297 | 133686965 | 44.3333333333333 | 0 | M37_down |
| *Zm00001eb153160* | chr3 | 202610008 | 202620083 | 44.6666666666667 | 0.666666666666667 | M37_down |
| *Zm00001eb167240* | chr4 | 7443264 | 7449070 | 67.6666666666667 | 8.33333333333333 | M37_down |
| *Zm00001eb169170* | chr4 | 17995258 | 18002847 | 264.666666666667 | 45.3333333333333 | M37_down |
| *Zm00001eb169180* | chr4 | 17999307 | 18001378 | 242.333333333333 | 44 | M37_down |
| *Zm00001eb170160* | chr4 | 23394257 | 23395270 | 101.666666666667 | 2.33333333333333 | M37_down |
| *Zm00001eb179100* | chr4 | 81667581 | 81680772 | 50 | 0.666666666666667 | M37_down |
| *Zm00001eb186550* | chr4 | 157767412 | 157769448 | 42.3333333333333 | 1 | M37_down |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Zm00001eb190720* | chr4 | 174233528 | 174235021 | 53.6666666666667 | 5.33333333333333 | M37_down |
| *Zm00001eb192830* | chr4 | 180653426 | 180658586 | 130.333333333333 | 0.666666666666667 | M37_down |
| *Zm00001eb201940* | chr4 | 217710718 | 217719116 | 173.333333333333 | 7.66666666666667 | M37_down |
| *Zm00001eb218610* | chr5 | 17198423 | 17202806 | 47.3333333333333 | 10 | M37_down |
| *Zm00001eb219710* | chr5 | 20985418 | 21032146 | 2079 | 171.333333333333 | M37_down |
| *Zm00001eb226760* | chr5 | 57788864 | 57790637 | 54 | 7.33333333333333 | M37_down |
| *Zm00001eb229190* | chr5 | 68086545 | 68088407 | 80.6666666666667 | 1 | M37_down |
| *Zm00001eb234750* | chr5 | 101619657 | 101621419 | 313.666666666667 | 93.3333333333333 | M37_down |
| *Zm00001eb234810* | chr5 | 101819233 | 101822718 | 44 | 0.666666666666667 | M37_down |
| *Zm00001eb241470* | chr5 | 166057998 | 166061646 | 43 | 0.333333333333333 | M37_down |
| *Zm00001eb260620* | chr6 | 8593290 | 8594835 | 113.666666666667 | 1.66666666666667 | M37_down |
| *Zm00001eb266140* | chr6 | 41352690 | 41353316 | 48 | 3.66666666666667 | M37_down |
| *Zm00001eb281150* | chr6 | 131544671 | 131545466 | 65 | 3 | M37_down |
| *Zm00001eb286570* | chr6 | 152786353 | 152795277 | 70.3333333333333 | 3.33333333333333 | M37_down |
| *Zm00001eb287500* | chr6 | 155980917 | 155985184 | 43 | 2 | M37_down |
| *Zm00001eb288580* | chr6 | 159457795 | 159468677 | 1653.33333333333 | 394.333333333333 | M37_down |
| *Zm00001eb297600* | chr6 | 179643623 | 179645636 | 80 | 0.666666666666667 | M37_down |
| *Zm00001eb321490* | chr7 | 155327162 | 155328812 | 55.3333333333333 | 0 | M37_down |
| *Zm00001eb322380* | chr7 | 160977249 | 160992924 | 877.666666666667 | 145.666666666667 | M37_down |
| *Zm00001eb341230* | chr8 | 46704359 | 46707233 | 495 | 9.66666666666667 | M37_down |
| *Zm00001eb355930* | chr8 | 137749111 | 137750559 | 183.333333333333 | 7.66666666666667 | M37_down |
| *Zm00001eb358660* | chr8 | 149012593 | 149027968 | 3829.66666666667 | 850.666666666667 | M37_down |
| *Zm00001eb360610* | chr8 | 155836951 | 155841244 | 2598.66666666667 | 667.666666666667 | M37_down |
| *Zm00001eb366470* | chr8 | 172214292 | 172220065 | 4367 | 879 | M37_down |
| *Zm00001eb369100* | chr8 | 177453257 | 177454219 | 325 | 3.33333333333333 | M37_down |
| *Zm00001eb373310* | chr9 | 9613369 | 9615299 | 82.6666666666667 | 0.333333333333333 | M37_down |
| *Zm00001eb373320* | chr9 | 9613394 | 9615299 | 82.6666666666667 | 0.333333333333333 | M37_down |
| *Zm00001eb374850* | chr9 | 15378295 | 15379012 | 66.3333333333333 | 4.66666666666667 | M37_down |
| *Zm00001eb382850* | chr9 | 62870918 | 62872776 | 101.666666666667 | 5.33333333333333 | M37_down |
| *Zm00001eb385100* | chr9 | 84052439 | 84059017 | 242.666666666667 | 69 | M37_down |
| *Zm00001eb390160* | chr9 | 116149068 | 116151841 | 185.333333333333 | 0.333333333333333 | M37_down |
| *Zm00001eb401320* | chr9 | 155897000 | 155899804 | 222 | 15 | M37_down |
| *Zm00001eb403700* | chr9 | 160090375 | 160092982 | 279 | 23.6666666666667 | M37_down |
| *Zm00001eb408630* | chr10 | 12861053 | 12867357 | 3916.66666666667 | 1050.66666666667 | M37_down |
| *Zm00001eb410610* | chr10 | 26020677 | 26023937 | 129.666666666667 | 1.33333333333333 | M37_down |
| *Zm00001eb419730* | chr10 | 102677792 | 102678850 | 1682.33333333333 | 1 | M37_down |
| *Zm00001eb431870* | chr10 | 146906613 | 146917198 | 232 | 30.3333333333333 | M37_down |

**Supp. Table 4**. List of the 19 candidate genes selected at BC2 in the crosses with the M162W inbred line. Status column indicates which inbred parental allele downregulates the other one, and downregulates alleles in all subsequent generations.

| *genes* | chr | start | end | B73_mean_expr | M162_mean_expr | status |
|---|---|---|---|---|---|---|
| *Zm00001eb062000* | chr1 | 298408180 | 298424189 | 159.666666666667 | 5.66666666666667 | M162_down |
| *Zm00001eb090480* | chr2 | 125575489 | 125578130 | 53 | 2 | M162_down |
| *Zm00001eb112530* | chr2 | 226951976 | 226957382 | 58 | 0.666666666666667 | M162_down |
| *Zm00001eb150160* | chr3 | 192128331 | 192133601 | 1237.33333333333 | 419.333333333333 | M162_down |
| *Zm00001eb209650* | chr4 | 248920663 | 248922782 | 162 | 33.3333333333333 | M162_down |
| *Zm00001eb223970* | chr5 | 39921823 | 39926221 | 314.333333333333 | 11.6666666666667 | M162_down |
| *Zm00001eb226250* | chr5 | 54637613 | 54642405 | 336.333333333333 | 2 | M162_down |
| *Zm00001eb240070* | chr5 | 158379379 | 158383359 | 57.3333333333333 | 0.333333333333333 | M162_down |
| *Zm00001eb271240* | chr6 | 86334802 | 86335710 | 138 | 6 | M162_down |
| *Zm00001eb271250* | chr6 | 86334829 | 86335367 | 138 | 6 | M162_down |
| *Zm00001eb277100* | chr6 | 113776057 | 113779256 | 52 | 0 | M162_down |
| *Zm00001eb307950* | chr7 | 65795536 | 65805003 | 57.6666666666667 | 1.33333333333333 | M162_down |
| *Zm00001eb366470* | chr8 | 172214292 | 172220065 | 4367 | 1399.33333333333 | M162_down |
| *Zm00001eb390160* | chr9 | 116149068 | 116151841 | 185.333333333333 | 0 | M162_down |
| *Zm00001eb419730* | chr10 | 102677792 | 102678850 | 1682.33333333333 | 1.33333333333333 | M162_down |
| *Zm00001eb012220* | chr1 | 39830879 | 39838448 | 377.333333333333 | 587.333333333333 | B73_down |
| *Zm00001eb095010* | chr2 | 162889727 | 162893463 | 24 | 57.3333333333333 | B73_down |
| *Zm00001eb157380* | chr3 | 216287975 | 216294074 | 212.333333333333 | 415.666666666667 | B73_down |
| *Zm00001eb345480* | chr8 | 80454829 | 80455179 | 15 | 74.6666666666667 | B73_down |

# Chapter 3

# Discovering long-range contacts involved in paramutation using Circular Chromosome Conformation Capture in maize

We designed a protocol using Circular Chromosome Conformation Capture (4C) to explore paramutation-associated long-range contacts. We used leaf and husk tissue from plants of the B73 genetic background (neutral *b* allele) and of the paramutagenic *B'* genetic background. We selected B' plants either as heterozygous or homozygous *mop1-1* mutants. We targeted the *b1*, *p1*, and *r1* TSS sites as well as a putative homolog of a rice *KEE (KNOT ENGAGED ELEMENTS*) *(* (Dong et al., 2018) and designed a bioinformatics pipeline to identify specific long-range contacts between alleles at the three loci we selected. Although preliminary, this work suggests more frequent long-range contacts in the *mop1* mutant than in the non-mutant contexts. However, few conclusions can be drawn regarding paramutation, as more work is needed to collect more conclusive data.

**My contribution**

I was fellowships recipient of two grants from the GAIA Doctoral School and the H2020 EU COST-action INDEPTH to visit Dr. Grob at the University of Zurich. I performed the experiments under his supervision, and bioinformatics and data analyses were performed together.

## Chapter 3 index

# Chapter 3

# Discovering long-range contacts involved in paramutation using Circular Chromosome Conformation Capture in maize

Juliette Aubert[1], Caroline Michaud[1], Julien Serret[1], Olivier Leblanc[1], Daniel Grimanelli[1], Stefan Grob[2]

[1] DIADE, University of Montpellier, CIRAD, IRD, Montpellier, France.

[2] Institute of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland.

## Introduction

Paramutation is a process sporadically observed in plants and animals by which gene expression pattern at specific loci is changed through the *trans*-silencing interactions between alleles. This *trans*-silencing is stable in subsequent generations, even after the original paramutagenic allele has been segregated away. This phenomenon was described at a few loci in maize, including *red1 (r1), pericarp color1 (p1), plant color1 (pl1),* and *booster1 (b1),* and it involves two types of alleles termed paramutagenic when they induce paramutation and paramutable when they are sensitive to paramutation. A known actor of paramutation is MEDIATOR OF PARAMUTATION (MOP1), a member of the RNA-directed DNA methylation (RdDM) pathway that is crucial to the biogenesis of 24-nt siRNAs. Homozygous *mop1-1/mop1-1* plants do not exhibit paramutation, however, it is restored in heterozygous *Mop1/mop1-1* plants (Dorweiler et al., 2000). Furthermore, enhancer sequences enabling paramutation were identified for the four known loci of paramutation, all located on the same chromosome as the gene they regulate (Kermicle et al., 1995; Sidorenko and Peterson, 2001; Stam et al., 2002; Wang et al., 2017). A known example of paramutation enhancers are seven tandem repeats located 100 kb upstream of the *b1* gene (called *b1TR*) that regulate the paramutation status of the *b1* gene Interestingly, the DNA sequence of the *b1TR* is identical at the paramutable (*B-I)* and paramutagenic (*B')* alleles but they differs for methylation levels that are higher at *B'* (Haring et al., 2010). Interestingly, the requirement of *b1TR* for paramutation at *b1* is not dependant on its localisation as plants carrying transgenes expressing a *b1TR* hairpin RNA recapitulated most of *b1* paramutagenicity regardless of transgene insertion site (Arteaga-Vazquez et al., 2010). Therefore, *b1TR* acts in *trans* through the production of 24-nucleotide small interfering RNAs (24-nt siRNAs) to regulate *b1* expression levels.

For many years, scientists hypothesized an important role of 3-dimensional (3D) chromosome conformation in global epigenetic mechanisms and gene regulation. However, the lack of suitable technology delayed its study until Chromosome Conformation Capture (3C) technology was described in yeast (Dekker et al., 2002) and opened a world of 3C-inspired technologies adapted to different organisms and applications (Hövel et al., 2012; de Wit and de Laat, 2012). These technologies enable to identify chromatin contacts, either between two known loci (3C and ChIP loop), one known locus and the rest of the genome (4C) or all contacts taking place in the genome (Hi-C) (de Wit and de Laat, 2012). A nuclear structure with frequent chromatin contacts called the *KNOT* was described in *A. thaliana* (Grob et al., 2014), *D. melanogaster* (Grob et al., 2014) and *Oryza sativa* (Dong et al., 2018). In *A. thaliana*, the *KNOT* is composed of ten *KNOT ENGAGED ELEMENTs* (*KEEs*) that are spread in the genome and rich in TEs that produce siRNAs. In *A. thaliana*, transgenes can be silenced by increased *KNOT* contact, which can cause a paramutation-like behavior of the transgene in subsequent generations (Grob and Grossniklaus, 2019). This observation questions the role of chromatin contacts and *KEEs* in paramutation. Interestingly, 3C experiments showed that the *b1TR* are involved in physical contacts with the *b1* transcription start site (TSS) in both *B'* and *B-I* epialleles. Interestingly, the chromatin state of this complex allows the formation of multiple loops involving several surrounding sequences of *b1* in the *B-I* epiallele, hence in a tissue- and epiallele-dependant manner (Louwers et al., 2009). The single loop formation depends on the presence of the *b1TR*, but the high *b1* expression depends on chromatin state, resulting in high *b1* expression in *B-I* and low in *B'*. This analysis demonstrated that chromatin contacts should be considered for their role in paramutation. So far, short-range contacts were identified as linked-to the *b1* paramutation, however we hypothesize that long-range contact are also involved in paramutation.

Based on the 3C experiment conducted in maize (Louwers et al., 2009), we designed a 4C protocol to detect contacts involving several key loci subjected to paramutation and a putative *KEE* in maize plants differing for their paramutation state. We also searched for chromatin contacts that are either specific or common to all the viewpoints selected, i.e. *b1TR*, *r1* TSS and direct repeats upstream of *p1*. Finally, we searched for contacts that occur in an allele-dependent manner. We identified high contact frequency of the putative *KEE* with centromeric sequences and we uncovered frequent and diverse contacts of the *p1* locus along chromosome 1, although this phenomenon is probably not related to paramutation. Interestingly, our results also suggest more frequent long-range chromatin contacts in the *mop1* mutant than in the other genetic backgrounds (WT). This work offers a preliminary insight into contacts associated with paramutation loci and a putative *KEE* in maize.

## Materials, methods & protocol design

### Plant Material

We selected the B73 inbred line that does not exhibit paramutation at the *b1* locus, as it possesses only one repeat from the *b1TR*, called b allele. It was provided by the Maize Genetics Cooperation Stock Center (University of Illinois, Urbana/Champaign, USA). We also selected both heterozygous *Mop1/mop1-1* and homozygous *mop1-1* mutants that were introgressed in a *B'* genetic background (Dorweiler et al., 2000) and were provided by V.L. Chandler (University of Arizona, Tucson, AZ, USA). The *B'* genetic background was not fully sequenced, however, the pBAC*B'1* that covers the *b1* region was sequenced (Stam et al., 2002). The heterozygous *Mop1/mop1-1* mutant is fully functional for the *b1* paramutation. However, the homozygous *mop1-1* mutant is not able to silence *b1* anymore and is hence a mutant of paramutation. Plants of the *b* allele (B73 inbreds), *Mop1/mop1-1*, and *mop1-1* genotypes were grown in a greenhouse at the French National Research Institute for Sustainable Development in Montpellier, France, with 14 hours day light (26°C during the day, 20°C at night). Tissue from the 8$^{th}$ leaf at 45 days post seeding (DPS) was collected in triplicate from three different plants for each genotype. We immediately proceeded with formaldehyde cross-linking to preserve chromatin contacts.

### Primer design and choice of restriction enzymes

The design of appropriate PCR primers to amplify contacts with loci of interest (viewpoints) is crucial for the quality of the 4C analysis (Grob, 2017). These primers should be designed facing outward as they should amplify the unknown DNA ligated to the viewpoint rather than amplify the viewpoint itself (**Fig. 1**). To identify paramutation-associated contacts, we designed primers around the *b1TR*, the *r1* TSS, and the repeats upstream (-6110 to -4842) *p1* TSS (Sidorenko and Peterson, 2001). It is worth noting that no paramutagenic activity is expected from the *r1* and *p1* loci, as we used the *B'* genetic background that is only paramutagenic for the *b1* locus, and contains neutral alleles at *r1* and *p1*. We used an R package (Stefan Grob; https://github.com/stefangrob/RprimeSuite) to design primers that are specific for the viewpoints in both *b* (*B73)* and *B'* (*Mop1/mop1-1* and *mop1-1*) genetic backgrounds. However, as no genomic sequence is available for the *B'* genetic background, the primers were designed using the B73 reference genome and verified manually for a perfect match using the pBAC*B'1* sequence. We generated primers allowing the use of the same restriction enzymes (RE), HindIII and DpnII, at all viewpoints.

Based on sequence homology with *A. thaliana* and high contact frequencies, several *KEEs* were identified in rice (Dong et al., 2018). Unfortunately, no *KEE* was identified in maize as yet, but

considering the high sequence similarity between rice and maize, we hypothesized that the maize genome likely contains homolog sequences to rice *KEEs*. To identify them, we selected conserved genetic motifs within rice and *A. thaliana KEEs* (Dong et al., 2018) and searched for sequence homology within the maize genome using the Blastn suite. The closest sequence was located on chromosome 3 (Chr3: 16,865,054-16,885,106) with 92.59% sequence similarity and $2.8^{e\ -11}$ e-value with the rice *KEE* (Chr4:2,315,000-2,315,500) (**Supp. Fig. 3**). In order to determine whether this sequence is a functional homolog of rice *KEEs*, we designed primers adequate for our 4C analysis, i.e. using the same REs as for the other viewpoints (HindIII and DpnII). It is important to note that this putative *KEE* cannot be validated as a true maize *KEE* until its contact with homolog sequences is demonstrated in maize. Therefore, we used 4C to identify loci with high contact frequency with this putative *KEE*. On the other hand, it was shown that increased *KNOT* contacts can cause paramutation-like gene expression (Grob and Grossniklaus, 2019). We therefore hypothesized a correlation between *KEE-b1* contact frequency and the level of expression of the *b1* gene.

## 4C experiment

We used a 4C protocol previously published by Stefan Grob (Grob, 2017). This involves initial cross-linking of chromatin contacts using formaldehyde, which we performed as described in the protocol, but we snap froze and stored the samples at -80°C. We then continued the protocol as described, with some adjustments to the NIB Buffer: we added 1 M DTT and Protease inhibitor tablets (Roche) and removed beta-mercaptoethanol. Samples were digested using the HindIII 6-cutter RE and ligated to enable binding of the cross-linked fragments together. Finally, removal of cross-linking and purification produced circular DNA that cover many (if not all) chromatin contacts, that are called the 3C templates (**Fig. 1a**). We assessed sample quality before and after the first digestion to assess the efficiency of the digestion and after the first ligation (the 3C template) to verify the ligation efficiency, using gel electrophoresis. Overall quality was sufficient to continue the 4C experiment because expected profiles were observed (**Fig. 1b**). To perform the 4C analysis, samples were digested using a second RE (4-cutter) (we used DpnII) and re-ligated. This ensures that circular DNA fragments are short and only contain one chromatin contact (i.e. two fragments). Efficiency of the primers was tested on the samples using PCR and gel electrophoresis, with an expected band of the size of the self-ligation product of the viewpoint fragment (delimited by the set of primers). Once the primers were validated, we ran 8 PCR per context to decrease PCR artefacts and we used the recommended components for Phusion™ High-Fidelity DNA Polymerase (ThermoFisher). We prepared the Illumina libraries using the KAPA LTP library prep Kit. Final PCR amplification was run with 3 cycles to avoid over-amplification. We used the Tapestation to validate library quality before sequencing and observed the expected profile

(**Fig 1c**). Libraries were sequenced using NextSeq550 machine at the CSHL Genome Center with paired-end reads.

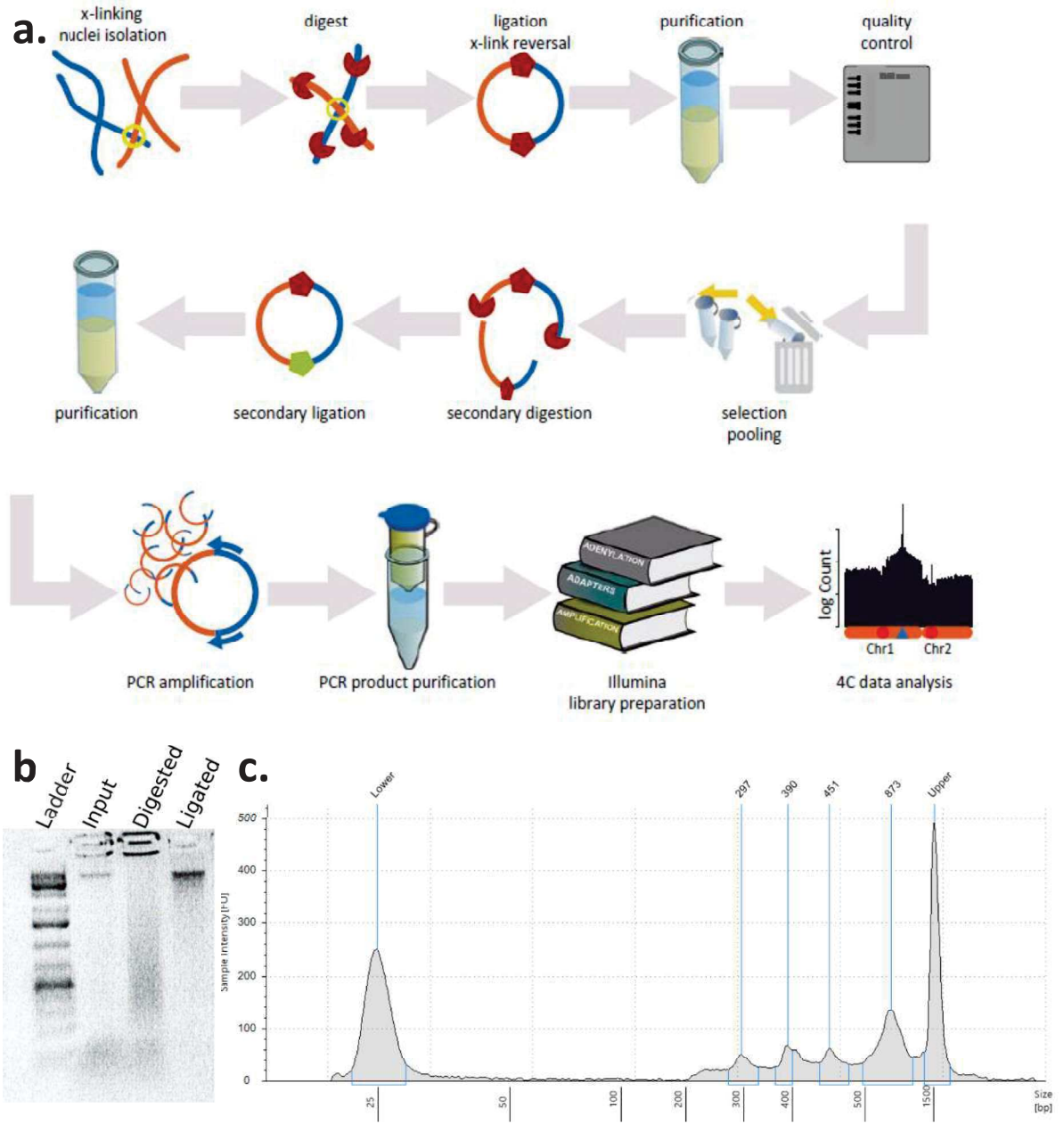



**Figure 1**. 4C experiment and its validation steps. **a)** The viewpoint is represented as the blue fragment. Image from (Grob, 2017). **b)** Evaluation of the 3C template quality. "Input" is the sample before digestion and shows one high weight band, as expected. "Digested" is migration of the samples after digestion and shows a smear of the expected size. "Ligated" is migration of the samples after ligation and purification and shows one high weight band as expected. 1 Kb ladder was used. **c)** Tapestation quality check of the 4C template before sequencing. The profile is as expected with various sizes of fragments and an increased molarity along fragment sizes.

### Realisation of a second 4C experiment

We conducted a second 4C analysis, for which we amplified the viewpoints using the following parameters: 98°C (30sec), 29 cycles of 98°C (15sec) – 58°C (30sec) – 72°C (45sec), 72°C (10 min). PCR products were pooled and purified using Monarch® PCR & DNA Cleanup Kit. A product length profile of each sample was validated using QIAxcel and the 4C libraries were prepared using NEBNext Ultra II DNA Library Prep Kit for Illumina, without size selection, and using AMPure XP Beads (NEB). Bioinformatic analysis will be initiated in the forthcoming weeks as we are still waiting for the sequencing raw data.

### Bioinformatics

Samples were processed following previously described protocol (Grob, 2017) by pooling the sequenced data per viewpoints, removing primer sequences, and aligning the trimmed fragments to the B73 genome version 5 (GenBank assembly accession: GCA_902167145.1). Total number of reads and number of aligned reads are described in (**Supp. Tab. 3**). These fragments correspond to loci interacting with the viewpoints. We binned the aligned fragments into 100 kb genomic windows and visualised viewpoints interactions along chromosomes using barplots for read count per window along chromosomes.

## Results and discussion

We first evaluated *intra*-chromosomal contacts of the four viewpoints in all studied genetic backgrounds by studying reads that aligned onto the same chromosome (**Fig. 2**). To validate the quality of the data, a high number of reads is expected in the surrounding area of the viewpoint. Both *p1* and *r1* showed a clear peak in all studied genetic background, indicating contacts of the viewpoint at a frequency that was not found to such levels in any other loci in the genome. We also added a blue triangle at each position where primers blast with > 95% identity, hence might amplify unwanted product. No peak was observed in the surrounding area of those blue triangle. Altogether, this validated that the respective primers were highly specific and only amplified regions interacting with the viewpoint. Interestingly, the *p1* peak in the surrounding area is wider than the one of *r1*, indicating that *p1* has more short-range contacts than *r1*. The region of contacts of the *p1* locus covers the entire arm of chromosome 1, making it difficult to identify a specific sequence that may cause such frequent intra-chromosomal contacts. The *KEE* viewpoint produced a few reads in the proximal region, but similar densities were also found at other loci in the genome (**Fig 4**). This suggests that the primers were not specific enough. On the other hand, *KEEs* are known for their unusual contact patterns in the organisms where they were described (Grob et al., 2014; Dong et al., 2018), and it is possible that the

putative maize *KEE* rarely interacts with proximal regions. Therefore, we proceeded with the *KEE* viewpoint analysis but regarded this data as preliminary requiring further confirmation. Finally, the *b1* viewpoint showed a weak peak for short-range contacts and higher peaks at many loci along the genome, especially around blue triangles (**Supp.Fig. 1**). This indicates that the primers were not specific enough and the data should not be considered for further analysis.

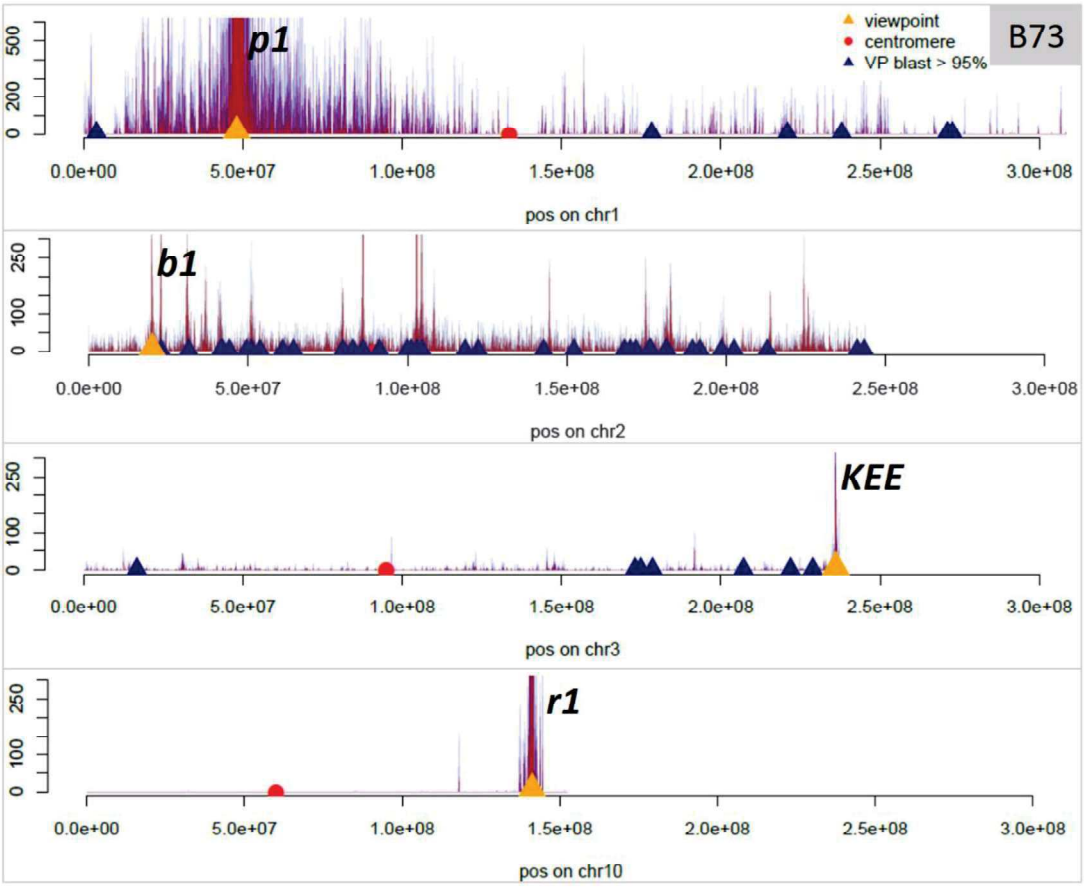



**Figure 2.** Short-range contacts of the four studied viewpoints in B73. Y axis is read counts. Orange triangle: position of the viewpoint; red circle: position of the centromere; blue triangles: positions where the viewpoints (VP) primers blast with more than 95% identity.

When focusing on the *p1* viewpoint, we found no other obvious peaks describing long-range interactions. It is however worth noting that short- and mid-range contacts are spread on the entire short arm of chromosome 1 with high peaks, suggesting high frequency of contacts with the *p1* locus (**Fig. 2**). When screening for *r1* contacts genome-wide, we detected three peaks present in all studied genetic backgrounds (although denser in the *mop1-1* mutant) on chromosomes 4, 5 and 8 (**Fig. 3A**). This suggests frequent contacts of the viewpoint with these regions. We also identified two loci interacting with the *r1* viewpoint on chromosome 1 in both *Mop1/mop1-1 and mop1-1* plants that are not observed in B73 plants (**Fig. 3B**). Both loci show clear peaks, indicating frequent contacts with *r1*

that are completely absent in B73. This difference may be linked to genetic background differences as both *Mop1/mop1-1* and *mop1-1* plants are of the *B'* genetic background, whereas B73 is a different genetic background altogether. Validating the presence of a gene of the locus in both genetic backgrounds by PCR would clarify whether this peak is linked to genetic differences.

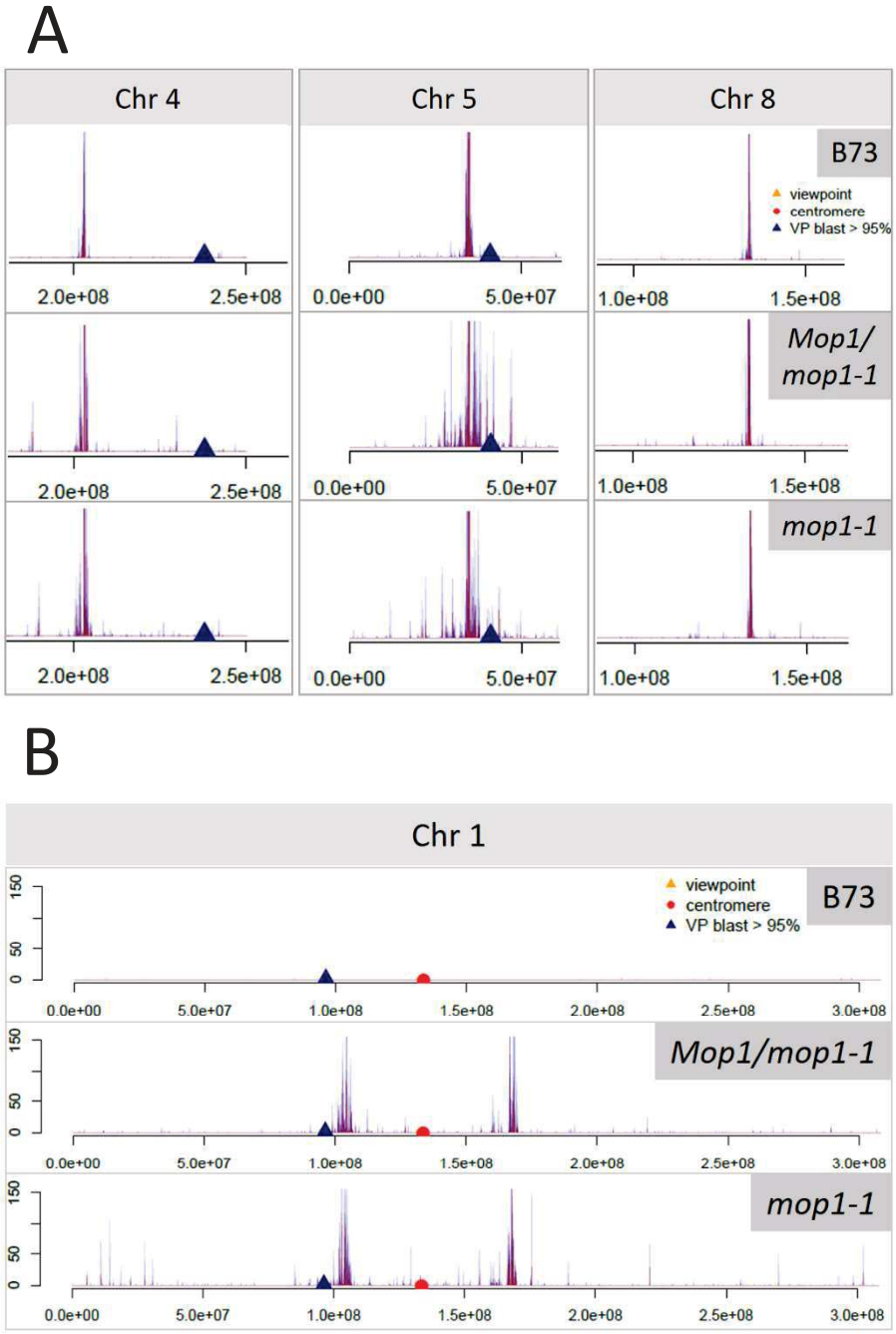



**Figure 3**. Contacts of the *r1* viewpoint on several chromosomes in all the studied genetic backgrounds. A) *r1* contacts with various chromosomes in all studied genetic backgrounds. Blue triangles are locations where the viewpoints primers blast with more than 95% identity. B) The two peaks observed on chromosome 1 in *Mop1/mop1-1* and *mop1-1* plants are not observed in the B73 genetic background.

Finally, we scanned the genome for *KEE* contacts and identified overall more frequent contacts of this viewpoint in *mop1-1* mutants than in the other genetic backgrounds. We also observed numerous peaks that were specific to the *mop1-1* mutant or showed a higher frequency in *mop1-1* than in the other genetic backgrounds (**Fig. 4A**). Interestingly, *KEE* peaks in the studied genetic backgrounds are often located near centromeres, suggesting that *KEEs* interacts preferentially with centromeres (**Fig. 4B**). Furthermore, we observed these contacts to centromeres at higher frequency in a *mop1-1* mutant context. Notably, *KEEs* were described as sequences containing TEs and producing 24-nt siRNA (Grob et al., 2014), which might explain a change of behaviour in a *mop1-1* mutant that produces reduced amounts of 24-nt siRNA (Nobuta et al., 2008). It was previously shown that there are additional *KEEs* in *A. thaliana*'s *met1* mutant (Grob et al., 2014), suggesting that demethylated *KEEs* after RdDM breakdown are more prone to participate in genome contacts. Note that these conclusions remain preliminary, as the short-range contacts of the *KEE* viewpoint detected using this data set are too weak to offer definite conclusions (**Fig. 2**). Furthermore, we observed higher frequency of long-range contacts in the *mop1-1* mutant than in the other genetic backgrounds for the four viewpoints. This may be linked to a weaker DNA methylation associated with reduced RdDM activity. This can be linked to a less compact chromatin that enables more long-range contacts. Notably, weakly methylated regions were found to correlate with higher chromatin contacts in various cell types of the human brain (Lee et al., 2019) and in the zebrafish (Yang et al., 2020). Similar behaviours were also described in *A. thaliana* where more frequent genome-wide contacts were observed in *ago4* mutants (Rowley et al., 2011; Böhmdorfer et al., 2016). All these results tend to indicate that DNA methylation, and more specifically RdDM, are essential for controlling chromatin contacts.

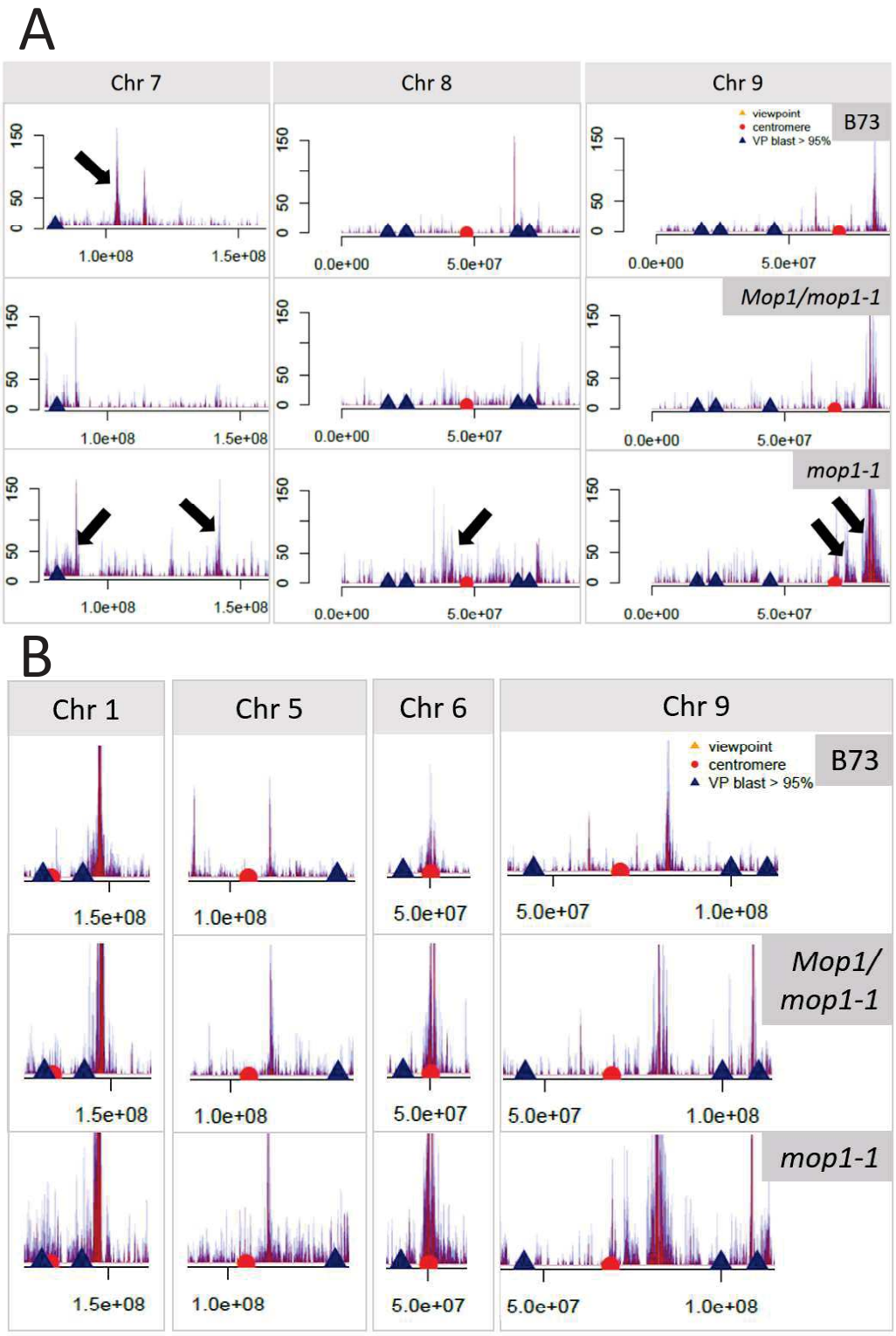



**Figure 4**. Contacts of the *KEE* viewpoint on portions of chromosomes. A) *KEE* contacts on chromosomes 7, 8 and 9. Black arrows point at peaks that are denser in one genetic background than in the other ones. B) *KEE* contacts with centromeres or in short-range regions of centromeres. The red circles indicate the approximate position of centromeres.

Overall, our results suggest that the *p1* repeats are involved in a high contact cluster on chromosome 1 short arm and that the *r1* TSS is involved in long-range contacts that vary depending on the genetic background. The *KEE* locus frequently interacts with centromeres, especially in a *mop1-1* mutant context thus suggesting that weaker DNA methylation is associated with more frequent and diverse chromatin contacts. Interestingly, this behaviour is similar to that of the *b1* paramutation locus as the *B'* allele shows higher methylation level than the *B-I* allele on the *b1TR* (Haring et al., 2010) and forms more loops with the *b1* TSS (Louwers et al., 2009).

## Limitations of the study

Despite drawing preliminary conclusions from our dataset, several issues concerning the experiments as well as the data analysis prevented further interpretation. We used the *mop1* mutant in the *B'* genetic background to allow studies of the long-range contacts associated in both a functional and a dysfunctional context for *b1* paramutation. However, we used leaf tissues that do not express the *b1* gene and do not show *b1* paramutation phenotypes. Therefore, it is unlikely that the chromatin contacts we detected reflect their paramutation state. Moreover, the primers designed around the *b1TR* were not specific enough and amplified many unwanted loci, which prevented further analysis. The two other loci of paramutation were correctly amplified but both viewpoints are of moderate interest as they are not paramutagenic in any of the three selected genetic backgrounds. Finally, we used the B73 reference genome to map all sequenced datasets, including those produced from *Mop1/mop1-1* and *mop1-1* plants, which are in the B' genetic background and, thus, significant sequence variation has to be expected. We observed chromatin contacts that were visible in one genetic background but not in the other, which can be interpreted either as a real novel contact or as an alignment artefact, caused by this sequence variation. Performing PCR analysis and sequencing of candidate regions in both genetic backgrounds would be required to solve this question.

This analysis focuses on identifying correlations between *B'* and *B-I* chromatin contacts and paramutation states. It is therefore necessary to use tissues with obvious changes in paramutation state between *B'* and *B-I* (i.e. husk tissues) and primers that display locus-specific chromatin contacts. We therefore designed a second 4C experiment that avoids most of these flaws. We selected the same genetic backgrounds (B73, *B' Mop1/mop1-1* and *B' mop1-1* ) but used husk tissue (leaf around the ear) that displays both paramutated and paramutable states depending on the background. Husk tissues were collected from immature ears (just before silk appearance) on three different plants for each genotype. We also used the sheath of six-week old *Mop1/mop1-1* plants that are interesting for evaluating the temporal aspects of the establishment of chromatin contact. Sheath tissues were

collected using three different *Mop1/mop1-1* plants. We focused our analysis on the *b1* locus and designed a set of primers at the *b1TR* and one at the *b1* TSS based on primers that were designed for a 3C analysis at the same loci (Louwers et al., 2009). We used Blastn to check for primer specificity in the B73 reference genome as well as on the pBAC*B'1* clone sequence (Stam et al., 2002). We also re-designed primers for the maize putative *KEE*. We applied the same protocol, except that we used two 4-cutter RE (DpnII and NlaIII) instead of a six-cutter and a four-cutter RE. This change will likely enable to improve the final resolution of short-range contacts. This analysis is still in progress as we are currently waiting to obtain the sequencing data.

## References


**Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K, Green PJ, Meyers BC, Chandler VL** (2010) RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proc Natl Acad Sci **107**: 12986–12991

**Böhmdorfer G, Sethuraman S, Rowley MJ, Krzyszton M, Rothi MH, Bouzit L, Wierzbicki AT** (2016) Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. Elife. doi: 10.7554/eLife.19092

**Dekker J, Rippe K, Dekker M, Kleckner N** (2002) Capturing chromosome conformation. Science (80- ) **295**: 1306–1311

**Dong Q, Li N, Li X, Yuan Z, Xie D, Wang X, Li J, Yu Y, Wang J, Ding B, et al** (2018) Genome-wide Hi-C analysis reveals extensive hierarchical chromatin interactions in rice. Plant J **94**: 1141–1156

**Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL** (2000) *mediator of paramutation1* Is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci. Plant Cell **12**: 2101–2118

**Grob S** (2017) Circular Chromosome Conformation Capture in Plants. Methods Mol Biol **1610**: 73–92

**Grob S, Grossniklaus U** (2019) Invasive DNA elements modify the nuclear architecture of their insertion site by KNOT-linked silencing in Arabidopsis thaliana. Genome Biol **20**: 120

**Grob S, Schmid MW, Grossniklaus U** (2014) Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila. Mol Cell **55**: 678–693

**Haring M, Bader R, Louwers M, Schwabe A, Van Driel R, Stam M** (2010) The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. Plant J **63**: 366–378

**Hövel I, Louwers M, Stam M** (2012) 3C Technologies in plants. Methods **58**: 204–211

**Kermicle JL, Eggleston WB, Alleman M** (1995) Organization of paramutagenicity in R-stippled maize. Genetics 141(1) 361-372

**Lee DS, Luo C, Zhou J, Chandran S, Rivkin A, Bartlett A, Nery JR, Fitzpatrick C, O'Connor C, Dixon JR,**

**et al** (2019) Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. Nat Methods **16**: 999–1006

**Louwers M, Bader R, Haring M, van Driel R, de Laat W, Stam M** (2009) Tissue- and Expression Level-Specific Chromatin Looping at Maize b1 Epialleles. Plant Cell Online **21**: 832–842

**Nobuta K, Lu C, Shrivastava R, Pillay M, De Paoli E, Accerbi M, Arteaga-Vazquez M, Sidorenko L, Jeong DH, Yen Y, et al** (2008) Distinct size distribution of endogenous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. Proc Natl Acad Sci U S A **105**: 14958–14963

**Rowley MJ, Avrutsky MI, Sifuentes CJ, Pereira L, Wierzbicki AT** (2011) Independent chromatin binding of ARGONAUTE4 and SPT5L/KTF1 mediates transcriptional gene silencing. PLoS Genet. doi: 10.1371/JOURNAL.PGEN.1002120

**Sidorenko L V., Peterson T** (2001) Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize p1 gene. Plant Cell **13**: 319–335

**Stam M, Belele C, Ramakrishna W, Dorweiler JE, Bennetzen JL, Chandler VL** (2002) The regulatory regions required for B' paramutation and expression are located far upstream of the maize b1 transcribed sequences. Genetics 162(2) 917-930

**Wang P-H, Wittmeyer KT, Lee T, Meyers BC, Chopra S** (2017) Overlapping RdDM and non-RdDM mechanisms work together to maintain somatic repression of a paramutagenic epiallele of maize pericarp color1. PLoS One **12**: e0187157

**de Wit E, de Laat W** (2012) A decade of 3C technologies: Insights into nuclear organization. Genes Dev **26**: 11–24

**Yang H, Luan Y, Liu T, Lee HJ, Fang L, Wang Y, Wang X, Zhang B, Jin Q, Ang KC, et al** (2020) A map of cis-regulatory elements and 3D genome structures in zebrafish. Nature **588**: 337–343

## Supplemental data

The following supplemental data are available for this chapter:

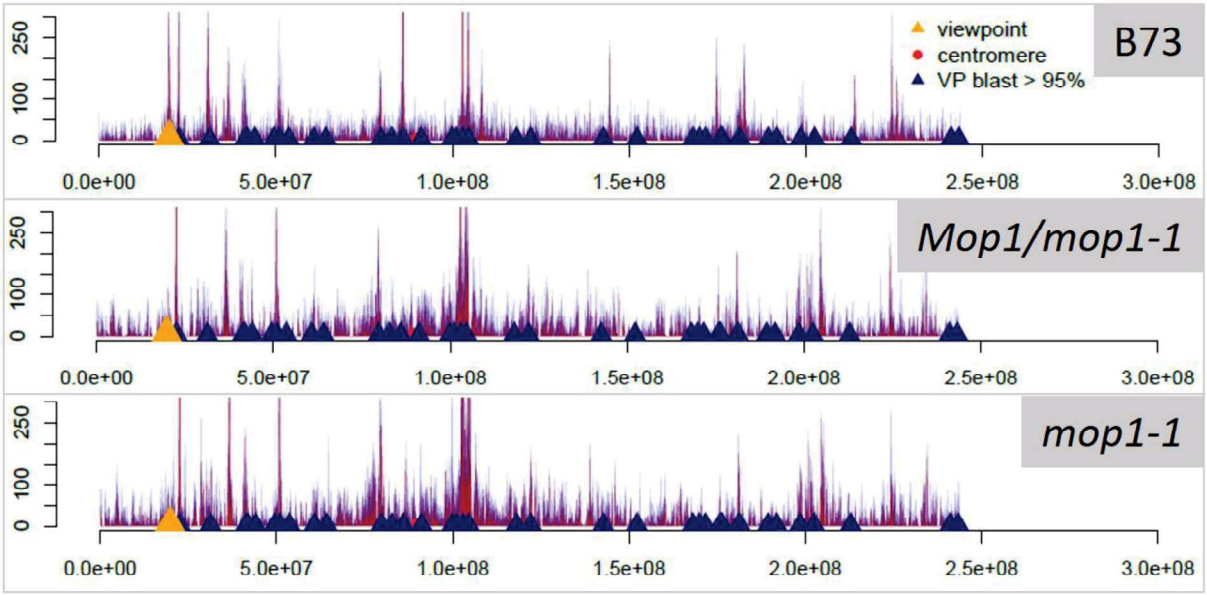
**Supplementary Figure 1.** Chromatin contacts of the b1TR on chromosome 2 in the three studied genetic backgrounds.

**Supplemental Table 1**. Primer sequences in the first 4C analysis.

**Supplemental Table 2**. Primer sequences in the second 4C analysis.

**Supplemental Table 3**. Identification of putative maize KEE in B73 Version 5 reference genome using sequence similarity with rice KEE.

**Supplemental Table 4**. Highest hits of the Blastn analysis to select a maize homolog of the rice conserved *KEE* sequence at position chr4:2315000-2315500 (Dong et al., 2018) in Osa1 Release 7.




**Supplementary Figure 1.** Chromatin contacts of the *b1TR* on chromosome 2 in the three studied genetic backgrounds. Numerous regions showing > 95% identity with the viewpoint primers (blue triangles) match to intense interaction peaks, indicating weak specificity of the primers to the viewpoint.

**Supplemental Table 1**. Primer sequences for the four viewpoints used in the first 4C experiment.

| Viewpoint | Primer sequence | Strand | Position in B73 genome (v5) | Size (bp) |
|---|---|---|---|---|
| *b1TR* | AACGA AATGC CGACC TCTAA GGCAC AAGCA CGAAA CA | + | chr2:19821257-19821293 | 246 |
| | GGCAC TATCT TGGCT TGGAA TGAT | - | chr2:19821070-19821047 | |
| *p1 repeats* | TGGTT AGTTC TTTAA TTCGA ACGAC T | + | chr1:47930450-47930475 | 462 |
| | TCGAC CCAAA ATATA TGCTC ATGTA CT | - | chr1:47930039-47930013 | |
| *r1* TSS | ACGTG CTAGC AGATG CTCAA ACT | + | chr10:141186624-141186646 | 493 |
| | GCCCT GGATT GGTTT TTATG CTCCA ATA | - | chr10:141186180-141186153 | |
| *KEE* | TCCTT TGCTT GAGGA CTTGG CGA | + | chr3:236067541-236067563 | 343 |
| | AAGGA CGGCA AGAAA AAGAA GGTAA AATCA A | - | chr3:236067250-236067220 | |

**Supplemental Table 2**. Primer sequences for the four viewpoints used in the second 4C experiment. Information relative to pBAC*B'1* is indicated when appropriate.

| Viewpoint | Primer sequence | Strand | Position in B73 genome (v5) | Size (bp) | Position in pBACB'1 | Size (bp) |
|---|---|---|---|---|---|---|
| *b1* TSS | CAAAG TGACC GAGCA AGACG | - | chr2:19822689-19822670 | 291 | 1275-1294 | 291 |
| | CGTTG CTTCT GGATG ACTGG | + | chr2:19822942 -19822961 | | 1003-1022 | |
| *b1TR* | GTGGT GACAG ACTCC TGGTC | - | chr2:19757226 -19757207 | 650 | 103718-103737 | 631 |
| | TGTCC TCAAA CTGAA AGCTT GC | + | chr2:19757836-19757857 | | 103087-103108 | |
| *KEE* | ACAAC AGCGA CGATG GTAGG | - | chr3:16877372-16877353 | 346 | | |
| | AATGC GACTT ACTGT GGGCA | + | chr3:16877680-16877699 | | | |

**Supplemental Table 3**. Summary for sequencing data in all viewpoints and genetic backgrounds. Mm: heterozygous *Mop1/mop1-1* plants in the *B'* genetic background; mm: homozygous *mop1-1* mutants in the *B'* genetic background; rep1: technical repeat number 1.

| | *r1* | | *p1* | | *b1* | | *KEE* | |
|---|---|---|---|---|---|---|---|---|
| | read number | aligned | read number | aligned | read number | aligned | read number | aligned |
| Mm_rep1 | 5941490 | 4E+06 | 2295542 | 740238 | 1313453 | 390180 | 2007542 | 1243622 |
| Mm_rep2 | 5006113 | 3E+06 | 2333378 | 929941 | 1083852 | 374178 | 1662049 | 1041947 |
| Mm_rep3 | 5260614 | 3E+06 | 2589543 | 1E+06 | 1461227 | 469390 | 1650743 | 972359 |
| mm_rep1 | 4819076 | 3E+06 | 1460106 | 536728 | 1224499 | 411461 | 1934897 | 1170798 |
| mm_rep2 | 5846536 | 4E+06 | 1854848 | 695241 | 1420425 | 518652 | 1482738 | 912087 |
| mm_rep3 | 5376003 | 3E+06 | 4791344 | 2E+06 | 1272741 | 456670 | 1040651 | 572123 |
| B73_rep1 | 4704422 | 4E+06 | 2780531 | 1E+06 | 1025145 | 352792 | 969991 | 865777 |
| B73_rep2 | 3746070 | 3E+06 | 1646467 | 664946 | 1141151 | 316308 | 932315 | 842388 |
| B73_rep3 | 4925397 | 4E+06 | 1945046 | 825527 | 1292439 | 352222 | 1459882 | 1342568 |

**Supplemental Table 4.** Highest hits of the Blastn analysis to select a maize homolog of the rice conserved KEE sequence at position chr4:2315000-2315500 (Dong et al., 2018) in Osa1 Release 7. The first hit was selected to design primers.

| Hit number | Position | e-value | % identity |
|---|---|---|---|
| 1 | Chr3:16865054-16885106 | 2.692e-11 | 92.59 |
| 2 | Chr6:20922316-20922388 | 9.683e-11 | 85.14 |
| 3 | Chr4:211510172-211510104 | 3.483e-10 | 85.71 |
| 4 | Chr2:176634850-176634782 | 3.483e-10 | 85.71 |
| 5 | Chr7:34328879-34328812 | 1.253e-9 | 85.51 |

# General discussion

## Discussion Index

## Contribution of modern sequencing methods

Paramutation was discovered in the 1950's in maize and intensively studied ever since, but the extent of the phenomena across living organisms and within genomes, the underlying regulatory mechanisms, and its emergence and conservation through evolution remain elusive (Brink, 1956; Chandler, 2010). One reason for that may be the inaccessible status of DNA and chromatin until recent advances in sequencing techniques. Most studies on paramutation concerned a few loci and mutants, which produced a great amount of knowledge on the mechanisms (Dorweiler et al., 2000; Stam et al., 2002; Hollick et al., 2005; Sidorenko et al., 2009). Next generation sequencing (NGS) technologies enable high-throughput DNA and RNA sequencing, hence of small RNA and 4C datasets as well, which allows to inventory genes and their regulation mechanisms. This offers massive advances in genetics and epigenetics by creating and processing significantly larger datasets, enabling the validation of previous hypotheses and the creation of new ones. This allows the widening of our views on paramutation in maize and approach its understanding with a new angle. This PhD work benefitted from these techniques that I implemented in the three chapters of my PhD thesis to better understand molecular and genetics mechanisms of paramutation.

## Results and limitations

### Chapter 1

In the first chapter, we investigated the mechanisms mediating paramutation: What happens to the *b1TR* 24-nt siRNAs after their biogenesis in the RdDM? Previously identified actors of paramutation are involved, at least in part, in the 24-nt siRNA biogenesis, and none were identified only in the effector complex of RdDM, probably due to a strong redundancy for the members of this complex (Dorweiler et al., 2000; Hollick et al., 2005; Sidorenko et al., 2009). Therefore, we asked whether the *b1TR* 24-nt siRNAs enable DNA methylation through the RdDM pathway or bypass it. To answer this question, we searched for RdDM effectors in maize and evaluated their involvement in paramutation. We first demonstrated the involvement of AGO104 in RdDM using Western blot, immunoprecipitation of AGO104 associated with small RNAs extraction, stem loop PCR and small RNAs sequencing. We showed that AGO104 is found in the same tissues and at the same cellular localisation than that of AGO9, its *A. thaliana* homolog (Olmedo-Monfil et al., 2010). Similarly to *At*AGGO9, AGO104 binds 24-nt siRNAs, including *b1TR* siRNAs. Then, we used a reverse genetics approach and showed that the *ago104-5* mutation disrupts paramutation at the *b1* locus, likely by impeding its establishment.

Considering the intermediate pigmentation phenotype we obtained in the reverse genetics population, we hypothesized that other ARGONAUTE proteins were probably involved in paramutation, therefore providing an explanation for the partial reversion of *B'* to *B-I* in the *ago104-5* mutant. Further research is needed to better characterize the role of AGO105 and AGO119, both putative homologs of *A. thaliana* AGO4 in the RdDM pathway and paramutation in maize (Singh et al., 2011). It is also possible that the *b1TR* 24-nt siRNAs bypass RdDM and use a different, yet unknown mechanism to mediate paramutation at *b1*. This will be discussed later, but shows that AGO104's involvement in paramutation does not close all questions on the mechanisms enabling paramutation. With this chapter, we showed that *ago104-5* causes a partial reversion of paramutation at *b1*, a phenotype that is transmitted through meiosis. However, it is not possible to conclude whether the ago104 mutation altered *b1* paramutagenicity as the paramutation state in progeny derived from a cross with *B-I* plants remains to be evaluated. This would allow to determine the effect of AGO104 on the establishment of the *b1* paramutagenicity and the occurrence of trans-generational effects.

Although much work remains to elucidate the mechanisms of paramutation, here we demonstrated that *b1TR* 24-nt siRNAs are acting through the RdDM effector complex. This suggests that these siRNAs are responsible for DNA methylation, a key step for enabling *b1* paramutation. Interestingly, both *B'* and *B-I* epiallele produce similar amounts of *b1TR* 24-nt siRNAs and, therefore, the factors responsible for DNA methylation increase in the *B'* context (Arteaga-Vazquez et al., 2010; Haring et al., 2010) remain to be elucidated.

## Chapter 2

In the second chapter, we addressed a question that has been barely explored: the role of paramutation in evolution. The emergence of paramutation, its biological interest and the number of genes regulated by this phenomenon are largely unknown. The only known organism using paramutation as a global silencing mechanism to protect its genome from foreign DNA is *C. elegans* (Sapetschnig et al., 2015). Here, I asked whether paramutation is a global silencing mechanism in maize. I selected genetically distant inbred lines and generated F1 and backcross generations. These materials allowed to search for loci subjected to paramutation by monitoring expression patterns genome wide. This approach uncovered 147 genes that exhibited a paramutation-like behaviour similar to that observed at the *b1* locus. They are all differentially expressed in the inbred lines, and silenced by the weakly expressed parent throughout backcrosses, even in the absence of the original weakly expressed allele. These 147 genes cover diverse biological functions and are distributed across

all chromosomes. They also show no difference in TE/repeats density and lacked overlap (4/147) with genes that are regulated by known actors of the RdDM pathway and DNA methylation.

I used a stringent protocol following the *b1* pattern of paramutation (i.e. strong, heritable silencing of paramutable alleles) allowing the use of statistical testing and, ultimately, avoiding the selection of false candidates. However, by using this model, I probably underestimated the number of genes subjected to paramutation. Indeed, it is likely that some genes were silenced in a heritable manner but were excluded from the selection because reduced expression level did not reach that of the lowly expressed parental gene. It is also possible that some genes are stably activated by the highly expressed parent, and stay activated stably through meiosis, similarly to *C. elegans* (reviewed in (Hollick, 2017).

One major unexpected result was the lack of overlap of my list of 147 candidate genes with the list of genes that are regulated by known actors of the RdDM pathway and methyltransferases. This conflicts with the dogma that paramutation is necessarily dependant of RdDM. This will be discussed later, but these results suggest that defining the paramutation phenomenon is a difficult task, with or without the use of biological mechanisms.

A previous study identified 145 paramutation candidates in B73xMo17 RILs. All were lowly expressed in the RILs, in similar manner to that of the low inbred parent, indicating a paramutation-like behaviour (Li et al., 2013). This number of genes is similar to the one I described in chapter 2, suggesting a conserved amount of paramutagenic alleles in maize inbred. However, we used a more distant inbred line, hence increasing the allelic variability, and we could have expected more candidates than in the B73xMo17 RILs. Furthermore, there are no genes in common between the ones I identified and the ones from Li et al. (2013), suggesting a high variation of loci subjected to paramutation between inbreds.

### Chapter 3

The study of chromatin contacts of KEE in *A. thaliana* (Grob and Grossniklaus, 2019) and of *b1* proximal regions (in both *B'* and *B-I*) (Louwers et al., 2009) showed a distinct connection between chromatin changes and paramutation states. It is however not known whether increased genome contacts can cause paramutation. Furthermore, only short-range contacts were studied to decipher paramutation, which might not fully cover paramutation properties (Louwers et al., 2009). In this last chapter, I designed and performed Circular Chromosome Conformation Capture (4C) in permissive and non

permissive genetic background for paramutation. Viewpoints on three identified loci of paramutation and a putative maize KEE were selected to determine putative chromatin contacts in mature leaves. I identified major chromatin contacts for the *p1* repeats (4 to 6 kb upstream of the TSS) with sequences located within its chromosome arm and that extended to the entire chromosome, however with lesser intensity. I also identified frequent chromatin contacts for the putative KEE locus with centromeric sequences on several chromosomes. Overall long-range chromatin contacts were more frequent in the *mop1-1* mutant, suggesting a link between weaker DNA methylation and higher chromatin density.

The frequent chromatin contacts of the *p1* viewpoint occur in both *mop1-1* and WT plants, suggesting no relationships with paramutation. Chromatin contacts of the *r1* TSS in the *B'* genetic background, but absent in the B73 inbred line were observed. This suggests either unspecific mapping on the reference genome or specific *r1* contacts taking place in both *mop1-1* and WT plants, hence not related to paramutation. This study included the *p1* and *r1* loci although it was conducted only in the *B'* paramutagenic background, which is not paramutagenic for these loci. We also conducted the study on the *b1* tandem repeats but, unfortunately, the weak specificity of the designed primers prevented the analysis of the data.

Although promising, these results are uninformative on paramutation. However, they reveal an interesting feature regarding chromatin, which shows overall more frequent long-range contacts in the *mop1-1* mutant. This mutant produces reduced amounts of 24-nt siRNA, hence partially inactivating the RdDM pathway and reducing the methyl deposit on CG, CHG and CHH contexts (Dorweiler et al., 2000). Weaker DNA methylation was recently associated with more frequent long-range chromatin contacts in several human brain tissues (Lee et al., 2019) and in zebra fish (Yang et al., 2020). Interestingly, a similar behaviour was described at the *b1TR* locus in maize, for which DNA methylation is higher and chromatin looping weaker at the *B'* epiallele than at the *B-I* epiallele (Louwers et al., 2009; Haring et al., 2010). Therefore, we can hypothesize that high levels of DNA methylation likely cause a decrease in chromatin contacts in maize.

## Mechanisms of paramutation

### What is the best definition of paramutation?

As described in the introduction, I chose to use a definition for paramutation that relies exclusively on allele *trans*-silencing and did not consider the underlying mechanisms. This includes the establishment and maintenance of paramutation when a paramutable allele is converted into a paramutagenic epiallele, and secondary paramutation when the new paramutagenic state is heritable (Chandler et al.,

2000). It is however worth noting that all identified mutants of paramutation in maize are members of RdDM (Dorweiler et al., 2000; Hollick et al., 2005; Sidorenko et al., 2009; Hollick, 2017), which demonstrates the importance of a fully functional RdDM pathway in paramutation. We therefore can question whether it makes sense to define paramutation using the 3 key steps of paramutation as defined by Brink (Brink, 1956) only (i.e. ignoring the RdDM pathway) while both criteria seem to be equally important to paramutation. As highlighted earlier, the RdDM pathway only exists in plants although paramutation was described in various eukaryotes outside the plant kingdom (Hollick, 2017; Erdmann and Picard, 2020). PiwiRNA were described as necessary to paramutation in animals, although these sRNAs are absent from the plant kingdom (Hollick, 2017), which indicates that distinct sRNA-dependent mechanisms are acting to mediate the allele *trans*-silencing behaviour required for establishing and maintaining paramutation. It is however worth noting that the plant RdDM pathway and the animal piwiRNA pathway are considered equivalent, as they both involve the production of siRNAs that bind to AGO proteins and enable methylation at specific DNA loci (Hövel et al., 2015). In regards, the choice I did when I initiated this work for a definition of paramutation remains valid for all eukaryotes as it does not consider the requirement for specific actors nor specific mechanisms.

In chapter 2, I monitored gene expression patterns in generations derived from genetically distant maize inbred lines to search for candidate alleles involved in paramutation. To achieve this, I used the key criteria of the mechanisms-free definition as guidelines to categorize expression patterns and select those matching a paramutation behaviour. Unexpectedly, the majority of the candidates did not match the list of genes regulated by members of RdDM in maize. This indicates that the RdDM pathway does not regulate the expression of the genes that we selected as candidates to paramutation. These candidates rigorously match the definition of paramutation in eukaryotes which suggests that RdDM is not the only mechanism that enables paramutation in maize. Animals are capable of paramutation in the absence of RdDM, indicating that both animals and plants use various mechanisms to stably *trans*-silence allele through meiosis (Hollick, 2017). We showed that *ago104* is required in the *b1* paramutation (see Chapter 1), but also showed that most candidates to paramutation do not require RdDM (see Chapter 2). We anticipate that the 147 new candidate genes to paramutation will facilitate the deciphering of the mechanisms driving paramutation.

### Crucial role of 24-nt siRNAs in paramutation

Although RdDM requirement was not considered for the paramutation definition, all described actors of paramutation in maize are RdDM members. Before this PhD thesis, all members of RdDM were, at least in part, involved in 24-nt siRNAs biogenesis that most likely enable paramutation by depositing

methyl modifications on cytosines at key loci (Dorweiler et al., 2000; Hollick et al., 2005; Sidorenko et al., 2009). Among examples is the production of 24-nt siRNAs from the *b1TR* that induces a higher *b1* expression (Stam et al., 2002). Chapter 1 describes ago104 as the first member of RdDM and actor of paramutation that do not participate in siRNA biogenesis. It binds 24-nt siRNAs and guides them to the POL V transcripts for silencing. This validates the requirement of the RdDM effector complex for paramutation at the *b1* locus, hence the entire pathway is used to enable paramutation. Moreover, *ago104* loss-of-function does not allow the full reversion of the paramutation phenotype, indicating that other AGO members are probably acting in the effector complex.

Interestingly, *ago104* escaped previous genetic screenings for paramutation phenotypes (Hollick and Chandler, 2001; Deans et al., 2020) and its mutation is not sufficient to fully revert the *B'* phenotype into a *B-I* phenotype. It is therefore possible that a second pathway is used by the 24-nt siRNAs, bypassing the effector complex of RdDM to enable paramutation. This would mean that the 24-nt siRNAs produced by the RdDM pathway are then capable to escape loading into AGO104 and act in other pathways to enable paramutation. This hypothesis provides an explanation for the results from Chapter 1 as 24-nt siRNAs bypass the effector complex of RdDM but are still produced by RdDM. This hypothesis however does not explain why, in Chapter 2, we identified 147 genes that are not dependant of MOP1 and MOP3, both involved in the biogenesis of 24-nt siRNA. It is possible that various exchanges between PTGS and RdDM enable 21-nt and 22-nt siRNA to be produced by PTGS and act in RdDM, similarly to the 24-nt siRNAs that are produced by RdDM and might act in PTGS (model described by (Cuerda-gil and Slotkin, 2016)). This model supports partially our results, as such bypasses are thought to occur too sporadically to create a paramutation behaviour.

Interestingly, the preliminary results from Chapter 3 indicate that chromatin contacts are stronger in an environment deprived of 24-nt siRNAs, such as the *mop1-1* mutant. This indicates that the 24-nt siRNAs are involved in the regulation of DNA conformation, most likely through DNA methylation like demonstrated in animals (Lee et al., 2019; Yang et al., 2020). The *mop1-1* mutant therefore exhibit alterations for chromatin contacts as well as for paramutation. Could paramutation and chromatin conformation be connected? This was strongly suggested by previous 3C analysis at the *b1* locus, although no causality could be demonstrated (Louwers et al., 2009). It is possible that siRNAs and DNA methylation regulate chromatin conformations, which in turn regulate paramutation. This would allow any mechanism linked to siRNAs or DNA methylation to indirectly regulate paramutation. This assumption therefore includes all organisms, whether they are capable of RdDM and DNA methylation or not. It is worth noting that the amount of 24-nt siRNAs produced by *B'* and *B-I b1TR* is identical,

therefore blurring the cause for differential methylation at *b1TR* and for different paramutation states (Arteaga-Vazquez et al., 2010).

## Paramutation in evolution

### Conservation of RdDM

A major axis of my thesis investigates the level of conservation of paramutation in evolution among eukaryotes. Paramutation was identified in plants and animals but requires different mechanisms since the RdDM pathway was never identified outside of the plant kingdom (Hollick, 2017; Erdmann and Picard, 2020). Most of the knowledge about RdDM comes from *A. thaliana*, but some conserved actors indicate that RdDM is somewhat conserved within the plant kingdom (Erdmann and Picard, 2020). In the first chapter of this thesis, I confirmed that maize *ago104* is an ortholog of *A. thaliana ago9* and plays the same role in RdDM. I also showed that it is involved in guiding the 24-nt siRNAs that enable the *b1* paramutation. Identification of this conserved member of RdDM between the *A. thaliana* dicot and maize monocot supports the high level of conservation of the RdDM pathway within the plant kingdom.

This however makes it harder to plead for the conservation of paramutation mechanisms within eukaryotes while a complete regulating pathway is absent from animals. Interestingly, the RdDM pathway shows similarities with the Post-Transcriptional Gene Silencing pathway (PTGS, also called RNAi pathway) that is found in all eukaryotes (**Fig. 1**) (Cuerda-gil and Slotkin, 2016; Erdmann and Picard, 2020). Both pathways require POLYMERASE, RNA DIRECTED RNA POLYMERASE, DICER LIKE and ARGONAUTE members and participate in the regulation of gene expression. It is therefore possible that this conserved pathway is involved in paramutation because of its close function similarity to plant RdDM. This leads to the hypothesis that there might be more than one mechanism capable of driving paramutation, although some mechanisms might be more efficient than others.

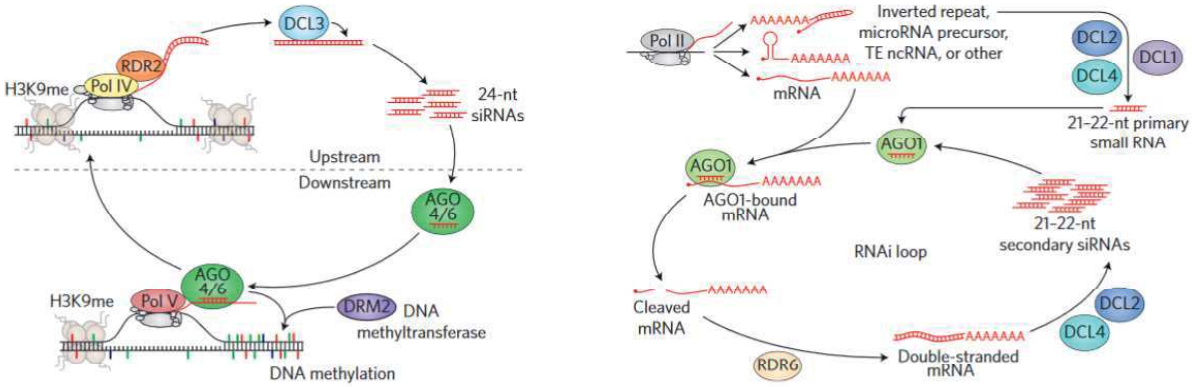



**Figure 1**. RNA directed gene silencing in plants. RdDM (left) and PTGS (right) both use the same families of actors, POLYMERASE, RNA DIRECTED RNA POLYMERASE, DICER LIKE and ARGONAUTE members. Image from (Cuerda-gil and Slotkin, 2016).

### Extent of loci subjected to paramutation in the maize genome

Gene expression in a hybrid progeny can be predicted for 80% of the expressed loci as the mid parent expression value (Stupar and Springer, 2006). Among the 20% remaining genes, expression patterns vary and can result from several phenomena, including: heterosis for which genes are expressed above the higher parental expression (Flint-Garcia et al., 2009) and imprinting for which allelic expression depends on the parental origin (Kermicle, 1969). Interestingly, 145 genes in B73xMo17 RILs followed the paramutation expected behaviour (Li et al., 2013). This number is notably similar to the number of genes I identified using a different approach (see Chapter 2), with 132 and 19 genes that followed the paramutation expected behaviour, indicating that distant inbred lines exhibit the same amount of paramutable alleles, although their genes do not overlap.

In total, 132 alleles are paramutable when we cross B73 with M37W, 19 alleles when we cross B73 with M162W, and 145 alleles for B73 with Mo17 RILs (Li et al., 2013). Only four alleles are in common between M37W and M162W crosses and are hence paramutable in both backgrounds, but the remaining alleles are paramutable only in the corresponding crosses. The fact that paramutagenicity of alleles is dependant on the inbred crossed suggests that their paramutagenicity occurred independently during inbred distancing, hence after *Zea mays* speciation. This theory supports the fact that the *r1* locus is not functional in teosinte and therefore most likely became paramutable when it became functional during or after *Zea mays* speciation (Hanson et al., 1996). Although there is no doubt that paramutation is a conserved mechanism among eukaryotes, both Li's (2013) and our analyses suggest that allele paramutagenicity is highly variable across the species scale. To understand the evolution of allele paramutagenicity, it would be interesting to evaluate gene expression through

generations in teosinte although such analysis with heterozygous teosinte lines can be arduous. Identification of alleles submitted to paramutation in teosinte would be interesting to compare to that of maize hybrids, in order to evaluate conservation of the phenomenon. Such analysis may also enable to identify ancestral loci submitted to paramutation in teosinte.

## Conclusions and perspectives

During my PhD research, I answered several crucial questions towards understanding the mechanisms and evolution of paramutation.

### Chapter 1

First, I described for the first time in maize a member of the RdDM pathway that is entirely devoted to the effector complex. I also showed that this effector, AGO104, is a regulator of paramutation at the *b1* locus. This shed new light on our understanding of the RdDM pathway in maize, as well as the mechanisms enabling paramutation. This partially answered my first research question on the possible role of the RdDM effector complex in paramutation in maize.

An experiment that could be performed rapidly is to validate the paramutagenicity of plants with an intermediate phenotype and a double heterozygous genotype (Mm;Aa) by crossing them with plants from the *B-I* genetic background. The pigmentation of the progeny should indicate whether plants with an intermediate phenotype can transmit paramutagenic alleles or whether they carry an unstable paramutation status. This result should contribute to refine the role of AGO104 in paramutation.

It could be informative to study further the small RNAs dataset that was extracted from AGO104 in several genetic backgrounds (B73 inbred and in the *B'* genetic background of *Mop1/mop1-1* and *mop1-1/mop1-1* plants). It gathers a list of targets of the RdDM through AGO104 in the entire genome. It could also inform on the span of action of AGO104 within RdDM.

Many effectors of the maize RdDM remain to be characterised, especially several ARGONAUTE proteins that likely play a similar role to AGO104, such as AGO105 and AGO119, both putative homologs of *A. thaliana* AGO4 (Singh et al., 2011). These proteins can be characterised using the same techniques as the ones I used in chapter 1.

Finally, a key information would be to determine if other small RNA-directed gene regulation pathways are involved in paramutation. To achieve this, one could identify a putative RDR6 in maize (a member of the PTGS pathway) and verify its involvement in paramutation.

### Chapter 2

To investigate the contribution of paramutation to evolutionary processes, I asked whether paramutation is a global silencing mechanism in maize. To answer this question, I identified 147 candidates expanding the number of loci subjected to paramutation far beyond the four hallmark loci in maize. The usual regulators of paramutation do not influence the level of expression of these 147 candidates, which opens new possibilities on the regulatory mechanisms of paramutation.

It would be interesting to investigate whether the *mop1-1* mutation affects the expression of 147 candidate genes by performing a RT-PCR on a *mop1-1* mutant in the B73 genetic background. We expect that candidate genes that are weakly expressed in B73 will be overexpressed in a *mop1-1* plant if they are regulated by MOP1. In a second step *mop1-1* mutant could be crossed to M37W or to M162W inbred to evaluate the level of expression of the 147 candidate genes. We expect that weakly expressed candidate genes will be overexpressed in the homozygous *mop1-1* mutant from the B73 genetic background if they are regulated by MOP1.

The identification of regulatory regions for the 147 candidate alleles is considered as an important step to identify a common feature that causes alleles to be paramutagenic. Surprisingly considering the similarities observed for *b1*, *p1*, *pl1*, and *r1* loci, we found no evidence for such feature in this work. An alternative approach could take advantage of the B73xM37W and B73xM162W NAM RILs populations (McMullen et al., 2009) to search for regions that are necessary to candidate alleles *trans*-silencing. The identification of RILS that express highly one of the candidate allele (that should be silenced if paramutagenicity is maintained) will enable to pinpoint regulatory regions for this candidate gene.

Finally, the description of 147 genes that are candidates to paramutation was a first step towards understanding the evolution of paramutation. A similar research of candidate to paramutation conducted in teosinte could be informative regarding the conservation of paramutation through evolution.

### Chapter 3

Finally, I generated preliminary results using 4C experiments that require further investigations. I identified a relationships between low 24-nt siRNAs amounts and more frequent long-range chromatin contacts. This highlights new possibilities to identify mechanisms enabling chromatin contacts, and perhaps paramutation. These results however do not answer my last research question: Can we identify long-range contacts associated with paramutation ?

The major task for this subject is to analyse the second dataset we produced. This analysis should bring information for long-range contacts occurring at the *b1TR* and the *b1* TSS. This analysis was conducted in B73 as well as in heterozygous and homozygous *mop1-1* mutants in the *B'* genetic background. Such dataset is interesting to decipher the chromatin contacts involved in paramutation.

A wider experiment could be to perform Hi-C in various tissues and various paramutagenic backgrounds. Such analysis would shed light onto the evolution of chromatin contacts in developing tissues as well as connect them to paramutation if possible. Possible genetic backgrounds to use for this Hi-C analysis are the B73 and M37W inbred lines that were used in Chapter 2. This would allow evaluating the chromatin contacts of the 147 candidate genes in both inbred lines.

Finally, although the results presented Chapter 3 are preliminary and require validation by the second ongoing 4C experiment, the long-range chromatin contact that are more frequent in the *mop1-1* mutant is puzzling and should be investigated in-depth. It would be interesting to determine what causes this phenomenon. If this is caused by methylation, any mutant of methylation should show more frequent long-range chromatin contacts using Hi-C experiments.

## General conclusion

With this three years PhD, I worked on identifying mechanisms of paramutation, as well as its occurrence in maize. I described AGO104 as an effector of the RdDM pathway in maize, and showed its importance in ensuring paramutation at the *b1* locus. I then created a hybrid population of distant inbred lines and identified a total of 147 genes that are candidates to paramutation, although they are not regulated by known actors of paramutation in maize. Finally I searched for long-range contacts associated with paramutation using 4C experiments, and identified a putative maize *KEE* as well as more frequent long-range contacts in the *mop1* mutant than in other genetic backgrounds.

## References


**Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K, Green PJ, Meyers BC, Chandler VL** (2010) RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proc Natl Acad Sci **107**: 12986–12991

**Brink RA** (1956) A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. Genetics **41**: 872–89

**Chandler VL** (2010) Paramutation ' s Properties and Puzzles. Science (80- ) **628**: 628–630

**Chandler VL, Eggleston WB, Dorweiler JE** (2000) Paramutation in maize. Plant Mol Biol **43**: 121–145

**Cuerda-gil D, Slotkin RK** (2016) Non-canonical RNA-directed DNA methylation. Nat Plants. doi: 10.1038/NPLANTS.2016.163

**Deans NC, Giacopelli BJ, Hollick JB** (2020) Locus-specific paramutation in Zea mays is maintained by a PICKLE-like chromodomain helicase DNA-binding 3 protein controlling development and male gametophyte function. PLoS Genet. doi: 10.1371/JOURNAL.PGEN.1009243

**Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL** (2000) *mediator of paramutation1* Is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci. Plant Cell **12**: 2101–2118

**Erdmann RM, Picard CL** (2020) RNA-directed DNA Methylation. PLOS Genet **16**: e1009034

**Flint-Garcia SA, Buckler ES, Tiffin P, Ersoz E, Springer NM** (2009) Heterosis Is Prevalent for Multiple Traits in Diverse Maize Germplasm. PLoS One. doi: 10.1371/journal.pone.0007433

**Grob S, Grossniklaus U** (2019) Invasive DNA elements modify the nuclear architecture of their insertion site by KNOT-linked silencing in Arabidopsis thaliana. Genome Biol **20**: 120

**Hanson MA, Gaut BS, Stec AO, Fuerstenberg SI, Goodman MM, Coe EH, Doebley JF** (1996) Evolution of anthocyanin biosynthesis in maize kernels: The role of regulatory and enzymatic loci. Genetics **143**: 1395–1407

**Haring M, Bader R, Louwers M, Schwabe A, Van Driel R, Stam M** (2010) The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. Plant J **63**: 366–378

**Hollick J, Chandler VL** (2001) Genetic Factors Required to Maintain Repression of a Paramutagenic Maize pl1 Allele. Genetics

**Hollick JB** (2017) Paramutation and related phenomena in diverse species. Nat Rev Genet **18**: 5–23

**Hollick JB, Kermicle JL, Parkinson SE** (2005) Rmr6 maintains meiotic inheritance of paramutant states in Zea mays. Genetics **171**: 725–740

**Hövel I, Pearson NA, Stam M** (2015) Cis-acting determinants of paramutation. Semin Cell Dev Biol **44**: 22–32

**Kermicle JL** (1969) Androgenesis conditioned by a mutation in maize. Science (80- ) **166**: 1422–1424

**Lee DS, Luo C, Zhou J, Chandran S, Rivkin A, Bartlett A, Nery JR, Fitzpatrick C, O'Connor C, Dixon JR, et al** (2019) Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. Nat Methods **16**: 999–1006

**Li L, Petsch K, Shimizu R, Liu S, Xu WW, Ying K, Yu J, Scanlon MJ, Schnable PS, Timmermans MCP, et al** (2013) Mendelian and Non-Mendelian Regulation of Gene Expression in Maize. PLoS Genet **9**: e1003202

**Louwers M, Bader R, Haring M, van Driel R, de Laat W, Stam M** (2009) Tissue- and Expression Level-Specific Chromatin Looping at Maize b1 Epialleles. Plant Cell Online **21**: 832–842

**McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, et al** (2009) Genetic properties of the maize nested association mapping population. Science (80- ) **325**: 737–740

**Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Arévalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP** (2010) Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature **464**: 628–632

**Sapetschnig A, Sarkies P, Lehrbach NJ, Miska EA** (2015) Tertiary siRNAs Mediate Paramutation in C. elegans. PLoS Genet **11**: 1005078

**Sidorenko L, Dorweiler JE, Cigan AM, Arteaga-Vazquez M, Vyas M, Kermicle J, Jurcin D, Brzeski J, Cai Y, Chandler VL** (2009) A dominant mutation in mediator of paramutation2, one of three second-largest subunits of a plant-specific RNA polymerase, disrupts multiple siRNA silencing processes. PLoS Genet. doi: 10.1371/journal.pgen.1000725

**Singh M, Goel S, Meeley R, Dantec C, Parrinello H, Michaud C, Leblanc O, Grimanelli D** (2011) Production of Viable Gametes without Meiosis in Maize Deficient for an ARGONAUTE Protein. Plant Cell **23**: 443–458

**Stam M, Belele C, Dorweiler JE, Chandler VL** (2002) Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev **16**: 1906–1918

**Stupar RM, Springer NM** (2006) Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics **173**: 2199–2210

**Yang H, Luan Y, Liu T, Lee HJ, Fang L, Wang Y, Wang X, Zhang B, Jin Q, Ang KC, et al** (2020) A map of cis-regulatory elements and 3D genome structures in zebrafish. Nature **588**: 337–343

# Résumé des objectifs et principaux résultats

*Génétique et épigénétique du phénomène de paramutation chez le maïs*

## Table des matières

## Introduction

Le cas de la paramutation est un phénomène rare qui implique un transfert mitotiquement et méiotiquement stable d'une information épigénétique d'une séquence homologue à une autre (Arteaga-Vazquez and Chandler, 2010; Giacopelli and Hollick, 2015). Il s'agit du trans-silencing d'un allèle fortement exprimé par un allèle faiblement exprimé. Ce trans-silencing est stable dans les générations suivantes, même en l'absence de l'allèle silencé d'origine. Certains loci de paramutation retournent spontanément à leur état fortement exprimé, mais d'autres, comme le locus *booster1* (*b1)*, n'en sont pas capables (Chandler et al., 2000). Fait intéressant, les gènes hautement et faiblement exprimés sont généralement génétiquement identiques et ne diffèrent que par le statut de la chromatine de certaines régions régulatrices, ce qui implique que ce phénomène soit régulé par des mécanismes épigénétiques (Sidorenko and Peterson, 2001; Stam et al., 2002a). La paramutation est établie dans les embryons en développement (Coe, 1966), mais la dynamique des mécanismes d'établissement et de maintien de la paramutation n'est pas encore entièrement comprise et se produit probablement dans les tissus végétatifs (Hollick et al., 1995; Chandler et al., 2000; Haring et al., 2010). Ce phénomène intrigant a été décrit dans un nombre réduit de loci chez les plantes et les animaux comme *M. musculus* (Rassoulzadegan et al., 2006), *C. elegans* (Sapetschnig et al., 2015) ou *D .melanogaster* (De Vanssay et al., 2012; Ciabrelli et al., 2017).

L'évènement de paramutation le mieux décrit a lieu au locus *b1* chez le maïs. Dans ce contexte, l'allèle faiblement exprimé *BOOSTER'* (*B'*) va engendrer une diminution stable de l'expression de l'allèle fortement exprimé *BOOSTER-INTENSE* (*B-I*). *B-I* est alors changé en un nouvel allèle *B'* qui sera à son tour capable de changer des allèles *B-I* en nouveaux *B'*, même en l'absence de l'allèle *B'* d'origine. Cet évènement de paramutation nécessite la présence de 7 répétitions en tandem 100 kb en amont de *b1* (*b1TR*). Ces *b1TR* sont impliqués dans le silencing de *b1* via leur transcription en petits ARNs interférents longs de 24 nucléotides (24nt siRNA), produits par le mécanisme de RdDM (RNA directed DNA Methylation). La voie RdDM se découpe en 2 parties principales : 1- la biogénèse des siRNAs et 2- le complexe effecteur, qui va permettre la mise en place et le maintien d'un silencing fort. Aujourd'hui, tous les acteurs connus de paramutation sont impliqués, au moins partiellement, dans la 1ère partie de la voie RdDM, c'est-à-dire dans la biogénèse des siRNA. **Notre première question est donc la suivante : existe-t-il des acteurs de paramutation impliqués dans la 2ème partie de la voie RdDM, la voie des effecteurs ?**

La paramutation a été décrite pour 4 loci différents chez le maïs : *booster1* (*b1)*, *red1 (r1), plant color1 (pl1),* et *pericarp color1 (p1)*. Ces loci ont tous été découverts grâce aux forts phénotypes qu'ils impliquent (pigmentation bordeaux de différents tissus de la plante). Une recherche plus vaste de

candidats à la paramutation a identifié 145 loci potentiellement paramutagéniques dans une population RILs entre B73 et Mo17 en utilisant du RNAseq (Li et al., 2013). **Notre seconde question est donc de déterminer si la paramutation est un mécanisme global de régulation de l'expression des gènes chez le maïs. Si c'est le cas, nous souhaitons identifier des caractéristiques communes à tous les candidats, pour déterminer ce qui cause la paramutation.**

Des travaux précédents ont montré que les *b1TR* sont impliqués dans la formation de boucles d'ADN avec différents locus en amont de *b1*. Ces boucles sont tissus-spécifiques et diffèrent en fonction de l'épi-allèle étudié (*B'* ou *B-I)* (Louwers et al., 2009). Ces zones d'interactions en *cis* sont potentiellement importantes pour la régulation de l'expression de *b1*. **Nous entreprenons de pousser plus loin cette question en cherchant des interactions en *trans* qui seraient impliquées dans la régulation de gènes paramutés chez le maïs, par l'utilisation de la méthode du 4C (Circular Chromosome Conformation Capture).**

## Questions de recherche et objectifs de la thèse

1. Quels sont les mécanismes impliqués dans le phénomène de paramutation ?
    - Qui sont les acteurs du complexe effecteur chez le maïs et sont-ils impliqués dans la paramutation ?

2. Les différentes formes de paramutation représentent-elles un (épi)phénomène, ou une forme plus globale de régulation de l'expression des gènes ?
    - Peut-on identifier des gènes subissant des évidences de trans-silencing dans des croisements entre lignées divergentes, et ces phénomènes de trans-silencing sont-ils méiotiquement stables ?
    - Peut-on identifier une caractéristique commune entre tous les candidats identifiés ?

3. Peut-on identifier des interactions entre chromosomes lors du phénomène de paramutation ?
    - Par la technique du 4C (Circular Chromatin Conformation Capture), peut-on mettre en évidence des interactions en 'trans' entre allèles impliqués dans la paramutation ? (En collaboration avec Stefan Grob, Université de Zurich)

## Quels sont les mécanismes impliqués dans le phénomène de paramutation ?

### Stratégie de recherche et moyens mis en œuvre

La voie RdDM est très bien décrite chez *A. thaliana*. La biogénèse des 24-nt siRNAs est assurée par le transcrit initial de POLYMERASE IV (POL IV), qui est pris en charge par RNA DEPENDANT RNA POLYMERASE 2 (RDR2) afin de créer un transcrit double brin (RDR2 est appelé MEDIATOR OF PARAMUTATION 1 ou MOP1 chez le maïs). DICER LIKE3 (DCL3) coupe ensuite ces transcrits en fragments doubles brins de 24 nucléotides : les siRNAs (Margis et al., 2006; Zhang et al., 2019). Les acteurs de cette première partie de RdDM ont été identifiés chez le maïs. En revanche, les membres de la seconde partie de la voie RdDM ne sont pas identifiés chez le maïs, uniquement chez *A. thaliana* où ARGONAUTE (AGO)4/6/9 prennent en charge les siRNAs afin de les guider jusqu'aux transcrits de POL V (Havecker et al., 2010; Olmedo-Monfil et al., 2010; Matzke et al., 2015). C'est ce complexe, combiné avec DOMAINS REARRANGED METHYLTRANSFERASE1 et 2 (DRM1 et 2), qui va permettre un silencing fort et durable du locus transcrit par POL V chez *A. thaliana* (Law and Jacobsen, 2010; Matzke et al., 2015; Zhang et al., 2019).

Pour déterminer si AGO104 est un effecteur de la voie RdDM, nous avons sélectionné le mutant *mop1-1/mop1-1* (*mm*) qui perturbe la RdDM en diminuant les quantités de 24nt siRNA. Les plantes hétérozygotes pour la mutation *mop1-1* (*Mm*) conservent une voie RdDM pleinement opérationnelle (Dorweiler et al., 2000; Zhang et al., 2019). Par homologies de séquences avec les acteurs de la RdDM chez *A. thaliana*, nous avons sélectionné AGO104, un orthologue probable de *AtAGO9* qui est potentiellement impliqué dans la voie des effecteurs de la RdDM pour guider les 24-nt siRNA jusqu'aux transcrits de POL V. Nous avons immunoprécipité AGO104 dans des épis immatures, et avons extrait ses petits ARN (small RNA-IP). Leur séquençage a permis d'identifier les loci qui expriment ces petits ARNs. Ces données ont été générées pour B73 et *mop1/mop1-1* (*Mm*), qui sont capables de paramutation, mais aussi chez *mop1-1/mop1-1* (*mm*) qui est incapable de paramutation. Ces jeux de données permettent de comparer le comportement de AGO104 dans des plantes capables ou incapables de paramutation. Par la suite, nous avons validé le rôle de AGO104 dans la paramutation en créant une population paramutagénique avec la mutation *KO ago104-5* afin d'évaluer la conservation des phénotypes de paramutation dans un fond mutant pour AGO104.

### Résultats

Nous avons d'abord sélectionné un anticorps anti AGO104 dont l'efficacité a été démontrée précédemment (Singh et al., 2011) afin d'immunoprécipiter AGO104 dans différents tissus. Après avoir

démontré que AGO104 est présent dans les mêmes tissus reproducteurs que AGO9 chez *A. thaliana* et a la même localisation cytoplasmique (via immunolocalization), nous avons extraits les petits ARN capturés par AGO104. Nous avons démontré que AGO104 est bien un membre de la RdDM chez le maïs en démontrant sa prise en charge de 24-nt siNRA qui sont RdDM-dependants. Nous avons ensuite montré que AGO104 prend en charge des petits ARN qui proviennent des *b1TR* (région régulatrice de *b1*) en séquençant les petits ARNs issus des IP de AGO104. Ceci suggère un rôle de AGO104 dans la prise en charge des petits ARNs impliqués dans le silencing de b1.

Afin de vérifier le rôle de AGO104 dans la paramutation au locus *b1*, nous avons sélectionné un mutant *ago104-5* que nous avons introgressé dans le fond génétique *B'* avec la mutation *mop1-1* hétérozygote (*B' Mm*) et homozygote (*B' mm*). La mutation *mop1-1* homozygote dans *B'* permet de valider le fonctionnement de la production d'anthocyanine en réponse à une surexpression de *b1*. Les plantes contrôles ont montré des pigmentations classiques de leur tige, avec des plantes *B' Mm* vertes et des plantes *B' mm* bordeaux. La présence de la mutation *ago104-5* dans *B' Mm* a déclenché l'apparition d'un phénotype nouveau, avec une pigmentation intermédiaire des tiges. Si 25 % d'entre elles portent un phénotype vert classique, 75% d'entre elles portent une pigmentation intermédiaire plus sombre que celle des plantes vertes *B' Mm*, mais plus claire que les plantes *B' mm*. Ce phénotype est transmis dans la descendance du rétrocroisement avec *B' mm*, même en l'absence de la mutation *ago104-5*. L'apparition de ce nouveau phénotype indique que AGO104 perturbe l'établissement de la paramutation au locus *b1*, et est donc un acteur de la paramutation.

## Conclusions

J'ai décrit pour la première fois chez le maïs un membre de la voie RdDM qui est entièrement consacré au complexe effecteur. J'ai également montré que cet effecteur, AGO104, est un régulateur de la paramutation au locus *b1*. Ceci a apporté un nouvel éclairage sur la compréhension de RdDM chez le maïs, ainsi que sur les mécanismes permettant la paramutation. Cela a répondu en partie à ma première question de recherche : qui sont les acteurs du complexe effecteur chez le maïs et sont-ils impliqués dans la paramutation ?

## Les différentes formes de paramutation représentent-elles un épiphénomène, ou une forme plus globale de régulation de l'expression des gènes ?

### Stratégie de recherche et moyens mis en œuvre

Trois étapes ont été décrites comme correspondant au schéma héréditaire du phénomène de paramutation, et ont permis de définir la paramutation. La première étape est appelée l'établissement de la paramutation, et se caractérise par l'établissement d'un trans-silencing stable d'un allèle sur un autre. La deuxième étape, le maintien de la paramutation, est la conservation de la faible expression de l'allèle sur plusieurs générations. Enfin, la dernière étape est appelée paramutation secondaire, car les allèles silencés peuvent à leur tour silencer de façon stable leurs homologues fortement exprimés, indépendamment de la présence de l'allèle silencé d'origine (Chandler et al., 2000). D'autres caractéristiques ont pu être ajoutées au fil des ans par de nombreux chercheurs sur ce qui définit un gène paramutable, par exemple des gènes génétiquement identiques dans les deux allèles et un niveau de méthylation identique dans les deux allèles, la production de 24-nt siRNA par la RdDM.... Cependant, ces caractéristiques ne s'appliquent pas à tous les exemples de paramutation chez les eucaryotes. J'ai donc choisi de sélectionner la première définition de la paramutation qui n'utilise que des caractéristiques génétiques et qui est la plus proche de ce qui a été décrit lors de la découverte du phénomène.

Les quatre loci de paramutation décrits chez le maïs impliquent de forts phénotypes de pigmentation sur divers tissus végétatifs et sur la graine. La transmission de ces phénotypes d'une manière non mendélienne est ce qui a attiré l'attention sur ces loci en premier lieu. Bien que ce comportement n'ait été décrit que pour quatre loci du maïs et semble être un phénomène isolé, un comportement similaire a été signalé dans de nombreux autres organismes, notamment Solanum lycopersicum (Gouil and Baulcombe, 2017), Mus musculus (Rassoulzadegan et al., 2006), Drosophila melanogaster (De Vanssay et al., 2012; Ciabrelli et al., 2017), Caenorhabditis elegans (Sapetschnig et al., 2015), et même chez l'homme (Bennett et al., 1997). Par conséquent, la paramutation ne peut pas être décrite comme isolée, car elle est hautement conservée chez les eucaryotes. Cependant, dans les organismes où la paramutation a été décrite, l'occurrence du phénomène reste limitée à un petit nombre de loci, sauf chez C. elegans qui utilise la paramutation comme un mécanisme de silencing global pour protéger son génome contre des séquences génomiques extérieures, notamment de piwiRNA transgéniques et de leurs séquences homologues (Sapetschnig et al., 2015). L'exemple de C. elegans est très intéressant car c'est le seul organisme connu capable à la fois de silencer et d'activer l'expression des gènes en

utilisant la paramutation. Il est possible que d'autres organismes utilisent la paramutation comme mécanisme de silencing global, mais cela n'a jamais été identifié par les études précédentes se concentrant sur des allèles isolés**.**

Nous avons sélectionné M37W et M162W, deux lignées génétiquement distantes de B73 afin de les croiser à B73 et d'étudier les interactions alléliques entre ces lignées au fur et à mesure des générations. Notre but est d'identifier des loci qui sont des candidats à la paramutation, c'est-à-dire qui sont différentiellement exprimés entre les lignées parentales, et qui sont ensuite aussi faiblement exprimés que le plus faible de deux parents, de manière constante au cours des rétrocroisements successifs. Ce comportement doit aussi être observé chez des plantes qui ne possèdent plus l'allèle parental faiblement exprimé : l'allèle restant doit à son tour silencer l'allèle fortement exprimé pour pouvoir être un candidat à la paramutation.

Nous avons séquencé les ARN messagers de la 4ème feuille des lignées sélectionnées (M37W, M162W et B73) et nous avons cherché le nombre de gènes différentiellement exprimés (DEG) par rapport à B73. Ceci nous a permis de valider que les lignées sont suffisamment distantes de B73 pour pouvoir conduire une étude sur la paramutation. Nous avons ensuite croisé M37W et M162W à B73 (F1), et leurs descendances ont été rétrocroisées deux fois avec chacun des parents (BC1 et BC2). L'ARN de chaque plante a été collecté sur la 4$^{ème}$ feuille et séquencé via illumina. Les données RNAseq au travers des générations nous permettent d'évaluer le nombre de gènes dont la transmission du silencing est méiotiquement stable.

## Résultats

Les données de RNA-seq des lignées parentales ont permis d'identifier entre 6 000 et 7 000 gènes différentiellement exprimés par rapport à B73. Tous ces gènes identifiés sont des cibles potentielles de trans-silencing à l'issue des croisements avec B73. Ces DEG sont ceux parmi lesquels nous avons cherché des candidats à la paramutation dans les générations suivantes. En utilisant les données RNAseq des générations suivantes, nous avons cherché des gènes dont la transmission du niveau d'expression évoquent ceux de la paramutation. Ils doivent être fortement exprimés chez un parent, faiblement exprimés chez l'autre parent, ainsi que faiblement exprimés dans toute la descendance, afin d'être retenus par notre crible. Nous avons ainsi identifié 132 gènes avec ce motif de transmission dans le BC2 de la population B73xM37W, ainsi que 19 gènes dans le BC2 de la population B73xM162W. Tous ces gènes sont trans-silencés de manière méiotiquement stable. Ils respectent les règles de la paramutation et sont donc d'excellents candidats à la paramutation. Pour répondre à la question de la

thèse: **Nous avons identifié des gènes subissant des évidences de trans-silencing dans des croisements entre écotypes divergents, et ces phénomènes de trans-silencing sont méiotiquement stables.**

Dans la littérature, les gènes identifiés comme paramutables (*b1, p1, pl1, r1*) ont comme seule similarité identifiée d'être impliqués dans la pigmentation de la plante. Parmi les gènes que nous avons identifiés ici, aucun d'entre eux n'est impliqué dans la pigmentation de la plante. Afin de mieux comprendre le rôle de la paramutation chez le maïs, nous souhaitons mettre en évidence une caractéristique commune, qui regroupe l'ensemble des gènes que nous avons identifiés comme candidats à la paramutation. Nous avons premièrement comparé la liste de gènes candidats à la paramutation avec la liste de gènes qui sont régulés par MOP1, par MOP3 (deux acteurs de la paramutation) ainsi que par ZMET2 et ZMET5 (deux méthyl-transférases qui visent les contextes CHG). Aucune superposition significative des listes n'a été observée, ce qui semble indiquer que ces acteurs ne régulent pas particulièrement nos candidats à la paramutation. Nous avons ensuite étudié la localisation des gènes candidats le long du génome, qui semblent équitablement répartis entre chromosomes, même si leur répartition au sein des chromosomes ne suit pas exactement celle des gènes exprimés. Enfin, la densité en éléments transposables en amont des gènes candidats n'est pas statistiquement différente de celle en amont de gènes sélectionnés aléatoirement. Ces différentes analyses n'ont donc pas permis d'identifier des caractéristiques communes entre tous les gènes candidats à la paramutation.

## Conclusions

Pour découvrir si la paramutation est impliquée dans les processus évolutifs, j'ai d'abord cherché si la paramutation est un mécanisme de silencing global. J'ai identifié 132 et 19 candidats à la paramutation en croisant des lignées distantes pendant plusieurs générations. Ceci démontre que la paramutation est plus commune que seulement quatre loci chez le maïs. Les régulateurs habituels de la paramutation n'influencent pas le niveau d'expression de ces candidats, ce qui ouvre de nouvelles possibilités sur les mécanismes qui régulent la paramutation.

## Peut-on identifier des interactions entre chromosomes lors du phénomène de paramutation?

### Stratégie de recherche et moyens mis en œuvre

La paramutation est définie comme un mécanisme de régulation en *trans* car c'est un phénomène ARN-dépendant (Stam et al., 2002b; Chandler, 2004). Ceci a été démontré au locus *b1*, en utilisant un transgène qui contient les 7 répétitions en tandem en amont de *b1* (*b1TR*) et qui exprime un ARN en épingle à cheveux. Celui-ci permet de recréer un allèle paramutagénique quel que soit le site d'insertion du transgène, ce qui démontre son fonctionnement en *trans* (Arteaga-Vazquez et al., 2010). De plus, les interactions en *cis* de *b1* sont corrélées au statut de paramutation de b1, ce qui indique que la conformation du génome et ses interactions sont probablement impliquées dans le contrôle de la paramutation (Louwers et al., 2009). L'utilisation de la technologie de capture de la conformation des chromosomes (3C) a permis de révéler une boucle de la chromatine entre les *b1TR*, le TSS de *b1* et les répétitions directes situées entre les deux, qui sont spécifiques du tissu et de l'allèle. Bien que cette conformation de la chromatine soit associée à la paramutation, il n'est pas clair si elle est la cause de la paramutation ou sa conséquence. En outre, la paramutation est connue pour être un mécanisme de trans-silencing et l'étude des seules interactions *cis* pourrait ne pas couvrir entièrement ses propriétés.

Nous avons cherché des interactions en *trans* par la technique 4C (Circular Chromosome Conformation Capture). Nous avons étudié les interactions associées à des locus clés de la paramutation chez le maïs, notamment les gènes *b1, r1,* et *p1*. Nous avons réalisé ces expériences sur des plantes capables de paramutation (B73 et *B' Mm*) ou non capables de paramutation (*B' mm)* afin de déterminer quelles sont les interactions qui sont impliquées dans les paramutations. Avec l'expertise de Stefan Grob (Université de Zurich), nous avons réalisé des 4C sur la 8$^{ème}$ feuille des plantes. Nous avons ensuite étudié l'occurrence des interactions sur chacun des chromosomes.

### Résultats

Sur cette analyse, le jeu de données des interactions de b1 est inutilisable à cause de la faible spécificité des amorces utilisées. Les interactions de *p1* et de *r1* montrent tous les deux des profils d'interactions en *cis* plus denses que leurs interactions en *trans*. Les interactions en cis de *p1* couvrent entièrement le bras court du chromosome 1 ce qui indique que les interactions de *p1* sont très denses sur ce chromosome. Aucune interaction n'est visible sur d'autres chromosomes, probablement car les interactions sur le chromosome 1 sont trop denses pour permettre de visualiser des interactions en

*trans*. Les interactions de *r1* sont différentes selon le fond génétique utilisé, ce qui peut indiquer des erreurs d'alignement ou bien des interactions allèles-spécifiques. Enfin, nous avons observé des interactions globalement plus denses dans les plantes mutantes pour *mop1*. Cela semble indiquer que la diminution de la quantité de 24-nt siRNA cause un dérèglement des interactions chromatiniennes. Ce changement est probablement dû à la baisse du niveau de méthylation global, lié à la diminution de la quantité de 24-nt siRNA dans le mutant *mop1-1*.

Nous avons conçu une deuxième expérience 4C en sélectionnant les trois mêmes fonds génétiques mais nous avons utilisé la feuille autour de l'épi qui présente l'avantage d'avoir soit un état paramuté, soit paramutable selon le fond génétique utilisé. Ce n'était pas le cas des feuilles utilisées précédemment, qui ne montrent aucun phénotype lié au statut de paramutation de *b1*. Nous avons concentré notre nouvelle analyse sur le locus *b1* et avons conçu un jeu d'amorces au niveau du *b1TR* et un autre au niveau du TSS de *b1*. Ces deux loci sont particulièrement intéressants car leurs interactions corrèlent avec le statut de paramutation de b1 (Louwers et al., 2009). Cette analyse est toujours en cours.

## Conclusions

En utilisant la 4C, j'ai produit des résultats préliminaires qui nécessitent une étude plus approfondie. J'ai identifié un lien entre des quantités plus faibles de 24-nt siRNA et des interactions chromatiniennes plus denses. Cela met en évidence de nouvelles possibilités de mécanismes permettant les interactions chromatiniennes, et peut-être la paramutation. Ces résultats ne répondent toutefois pas à ma dernière question de recherche : Pouvons-nous identifier des *trans*-interactions associées à la paramutation ?

La tâche principale restante sur ce sujet est d'analyser le second jeu de données que nous avons produit, afin de répondre à la question de recherche. Cette analyse devrait apporter des informations sur les trans-interactions des *b1TR* et du TSS de *b1* dans le génome. Cette analyse a été réalisée dans B73 ainsi que dans des mutants *mop1-1* hétérozygotes et homozygotes dans le fond génétique *B'*. Ce jeu de données est intéressant pour déchiffrer les interactions chromatiniennes impliquées dans la paramutation.

# Conclusion générale

Au cours de ce doctorat de trois ans, j'ai travaillé sur l'identification des mécanismes de paramutation, ainsi que sur son occurrence chez le maïs. J'ai décrit AGO104 comme un effecteur de la voie RdDM

chez le maïs, et montré son importance pour assurer la paramutation au locus *b1*. J'ai ensuite créé une population hybride de lignées distantes et identifié un total de 147 gènes candidats à la paramutation, bien qu'ils ne soient pas régulés par des acteurs connus de la paramutation chez le maïs. Enfin, j'ai recherché des contacts de longue portée associés à la paramutation en utilisant la 4C, et j'ai identifié un *KEE* potentiel du maïs. J'ai aussi identifié des contacts de longue portée plus fréquents chez le mutant *mop1* que dans d'autres fonds génétiques.

## Références


**Arteaga-Vazquez M, Sidorenko L, Rabanal FA, Shrivistava R, Nobuta K, Green PJ, Meyers BC, Chandler VL** (2010) RNA-mediated trans-communication can establish paramutation at the b1 locus in maize. Proc Natl Acad Sci **107**: 12986–12991

**Arteaga-Vazquez MA, Chandler VL** (2010) Paramutation in maize: RNA mediated trans-generational gene silencing. Curr Opin Genet Dev **20**: 156–163

**Bennett ST, Wilson AJ, Esposito L, Bouzekri N, Undlien DE, Cucca F, Nisticò L, Buzzetti R, Bosi E, Pociot F, et al** (1997) Insulin VNTR allele-specific effect in type 1 diabetes depends on identity of untransmitted paternal allele. Nat Genet **17**: 350–352

**Chandler VL** (2004) Poetry of b1 paramutation: cis- and trans-chromatin communication. Cold Spring Harb. Symp. Quant. Biol. pp 355–361

**Chandler VL, Eggleston WB, Dorweiler JE** (2000) Paramutation in maize. Plant Mol Biol **43**: 121–145

**Ciabrelli F, Comoglio F, Fellous S, Bonev B, Ninova M, Szabo Q, Xuéreb A, Klopp C, Aravin A, Paro R, et al** (2017) Stable Polycomb-dependent transgenerational inheritance of chromatin states in Drosophila. Nat Genet **49**: 876–886

**Coe EH** (1966) THE PROPERTIES, ORIGIN, AND MECHANISM OF CONVERSION-TYPE INHERITANCE AT THE *B* LOCUS IN MAIZE. Genetics **53**: 1035–1063

**Dorweiler JE, Carey CC, Kubo KM, Hollick JB, Kermicle JL, Chandler VL** (2000) *mediator of paramutation1* Is Required for Establishment and Maintenance of Paramutation at Multiple Maize Loci. Plant Cell **12**: 2101–2118

**Giacopelli BJ, Hollick JB** (2015) Trans-Homolog interactions facilitating paramutation in maize. Plant Physiol **168**: 1226–1236

**Gouil Q, Baulcombe DC** (2017) Frequent paramutation-like features of natural epialleles in tomato. BMC Genomics **19**: 1–15

**Haring M, Bader R, Louwers M, Schwabe A, Van Driel R, Stam M** (2010) The role of DNA methylation, nucleosome occupancy and histone modifications in paramutation. Plant J **63**: 366–378

**Havecker ER, Wallbridge LM, Hardcastle TJ, Bush MS, Kelly KA, Dunn RM, Schwach F, Doonan JH,**

**Baulcombe DC** (2010) The arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci. Plant Cell **22**: 321–334

**Hollick JB, Patterson GI, Coe EH, Cone KC, Chandler VL** (1995) Allelic interactions heritably alter the activity of a metastable maize pl allele. Genetics **141**: 709–719

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

**Li L, Petsch K, Shimizu R, Liu S, Xu WW, Ying K, Yu J, Scanlon MJ, Schnable PS, Timmermans MCP, et al** (2013) Mendelian and Non-Mendelian Regulation of Gene Expression in Maize. PLoS Genet **9**: e1003202

**Louwers M, Bader R, Haring M, van Driel R, de Laat W, Stam M** (2009) Tissue- and Expression Level-Specific Chromatin Looping at Maize b1 Epialleles. Plant Cell Online **21**: 832–842

**Margis R, Fusaro AF, Smith NA, Curtin SJ, Watson JM, Finnegan EJ, Waterhouse PM** (2006) The evolution and diversification of Dicers in plants. FEBS Lett **580**: 2442–2450

**Matzke MA, Kanno T, Matzke AJM** (2015) RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. Annu Rev Plant Biol **66**: 243–267

**Olmedo-Monfil V, Durán-Figueroa N, Arteaga-Vázquez M, Demesa-Arévalo E, Autran D, Grimanelli D, Slotkin RK, Martienssen RA, Vielle-Calzada JP** (2010) Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature **464**: 628–632

**Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I, Cuzin F** (2006) RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. doi: 10.1038/nature04674

**Sapetschnig A, Sarkies P, Lehrbach NJ, Miska EA** (2015) Tertiary siRNAs Mediate Paramutation in C. elegans. PLoS Genet **11**: 1005078

**Sidorenko L V., Peterson T** (2001) Transgene-induced silencing identifies sequences involved in the establishment of paramutation of the maize p1 gene. Plant Cell **13**: 319–335

**Singh M, Goel S, Meeley R, Dantec C, Parrinello H, Michaud C, Leblanc O, Grimanelli D** (2011) Production of Viable Gametes without Meiosis in Maize Deficient for an ARGONAUTE Protein. Plant Cell **23**: 443–458

**Stam M, Belele C, Dorweiler JE, Chandler VL** (2002a) Differential chromatin structure within a tandem array 100 kb upstream of the maize b1 locus is associated with paramutation. Genes Dev **16**: 1906–1918

**Stam M, Belele C, Ramakrishna W, Dorweiler JE, Bennetzen JL, Chandler VL** (2002b) The regulatory regions required for B' paramutation and expression are located far upstream of the maize b1 transcribed sequences. Genetics 162(2) 917-930

**De Vanssay A, Bougé AL, Boivin A, Hermant C, Teysset L, Delmarre V, Antoniewski C, Ronsseray S**

(2012) Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature **490**: 112–115

**Zhang Z, Teotia S, Tang J, Tang G** (2019) Perspectives on microRNAs and phased small interfering RNAs in maize (Zea mays L.): Functions and big impact on agronomic traits enhancement. Plants. doi: 10.3390/plants8060170