Technische Universität München
TUM School of Life Sciences

DNA methylation remodeling in F1 hybrids and its impact on heterosis

Ioanna Kakoulidou

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften

(Dr.rer.nat)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Aurélien Tellier

Prüfer der Dissertation:

1. Prof. Dr. Frank Johannes
2. Prof. Dr. Claude Becker

Die Dissertation wurde am 02.05.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 04.10.2023 angenommen.

Contents

Abstract **iii**

Zusammenfassung **v**

1 Introduction **1**
- 1.1 DNA methylation 2
- 1.2 Classification of DNA methylation remodeling events 3
- 1.3 Mechanisms of DNA methylation remodeling 5
  - 1.3.1 DMR-gain events 5
  - 1.3.2 DMR-loss events 8
  - 1.3.3 SMR-gain and loss events 8
- 1.4 Links to non-additive gene expression and heterosis 8
- 1.5 Outstanding questions in the field 9
- 1.6 Epigenetic Recombinant Inbred Lines (epiRILs) 9
  - 1.6.1 The *met1-3* epiRILs 10
  - 1.6.2 The *ddm1-2* epiRILs 10
  - 1.6.3 Epigenetic Hybrids (epiHybrids) 11
- 1.7 Epigenetic diversity as a predictive marker of hybrid's performance 13
- 1.8 Aim of the project 14

2 Materials and Methods **17**
- 2.1 Plant material 17
- 2.2 Plant cultivation and crosses 18
- 2.3 High-throughput phenotyping 18
- 2.4 Imaging and image analysis 19
- 2.5 Quantitative phenotypic analysis of heterosis 20
- 2.6 DNA and RNA sample preparation 20
- 2.7 Characterising hybrid remodeling 21
- 2.8 RNA-seq analysis 22
- 2.9 sRNA-seq analysis 22
- 2.10 DNA-seq analysis 23
- 2.11 DMR detection between the msCol plants and the Col-wt 24
- 2.12 Construction of an epiRIL linkage map 24

2.13 Annotation enrichment analysis . . . 25
2.14 Epigenome-wide Association Study (EWAS) . . . 25
2.15 Statistical, Graphics, and Visual Analyses . . . 25
2.16 Data availability . . . 25

3 Results **27**
3.1 Construction of a large epiHybrid population . . . 27
3.2 Patterns of local methylome remodelling in epiHybrids . . . 32
3.3 Distal regions undergo widespread co-remodeling in epiHybrids . . . 33
3.4 Parental DMRs direct methylome remodeling locally and at distal regions . . . 37
3.5 Parental pericentromeric DMRs trigger heterosis . . . 42
3.6 Phenotypically relevant *de novo* NADs . . . 49

4 Discussion **55**

References **63**

Acknowledgments **73**

Abstract

Heterosis is the superior phenotypic performance of F1 hybrids relative to their parents and it is influenced by genetic and epigenetic factors. Differentiating between these factors is challenging as parental lines used in hybrids have varying genetic and epigenetic profiles. A key challenge is studying the connection between parental and hybrid epigenomes in isolation from genetic variation to attribute phenotypic changes to epigenetic states. Recent research shows that DNA methylomes of F1 hybrids undergo substantial remodeling relative to their parents, but the extent to which this remodeling is attributed to specific differentially methylated regions (DMRs) in the parental genomes, and their contribution to phenotypic heterosis, remains poorly understood.

In this study, we generated 500 epigenetic hybrid (epiHybrids) families by crossing a male sterile plant with 500 different *ddm1*-derived epigenetic recombinant inbred lines (epiRILs). We conducted methylome, transcriptome, and phenotypic profiling for a number of parent-hybrid trios and our data revealed that hybrid methylomes underwent extensive remodeling. These remodeling events in hybrids' methylome were not isolated, but rather they substantially co-occurred among distant regions. Although the precise regulatory mechanisms are not yet fully understood, our data suggest that trans-acting 24nt small RNA (sRNA) plays a key role in this process. Importantly, we demonstrated that stably segregating parental pericentromeric DMRs facilitated methylation changes in the hybrids not only locally but also at distal regions, affecting thousands of target sites throughout the genome. Remarkably, many of these trans-induced changes were associated with the expression of nearby genes and were significantly associated with phenotypic heterosis.

Our research findings suggest a compelling model in which parental differences in DNA methylation patterns, specifically in the heterochromatin-rich pericentromeric regions, play a major role in the reorganization of hybrid methylomes and transcriptomes. These changes contribute to non-additive phenotypic effects that underlines the phenomenon of heterosis. The fact that these effects occur in isogenic plant material highlights the importance of epigenetic variation as a significant molecular determinant of this classical phenomenon.

Zusammenfassung

Die Heterosis beschreibt die überlegene phänotypische Leistung von F1-Hybriden im Vergleich zu ihren Eltern und beruht auf einer Kombination aus genetischen und epigenetischen Faktoren. Die Unterscheidung zwischen diesen Faktoren ist schwierig da die in Hybriden verwendeten Elternlinien unterschiedliche genetische und epigenetische Profile aufweisen. Eine zentrale Herausforderung ist die Untersuchung der Verbindung zwischen den Epigenomen der Eltern und der Hybride unabhängig von der genetischen zu untersuchen, um phänotypische Veränderungen auf epigenetische Zustände zurückführen zu können. Jüngste Forschungen zeigen, dass DNA Methylome von F1-Hybriden im Vergleich zu ihren Eltern erheblich umgestaltet werden, aber das Ausmaß aber das Ausmaß, in dem diese Umgestaltung auf spezifische differentiell methylierte Regionen (DMRs) in den und ihr Beitrag zur phänotypischen Heterosis ist jedoch noch wenig bekannt.

In dieser Studie erzeugten wir 500 epigenetische Hybridfamilien (epiHybrids), indem wir eine männlich sterile Pflanze mit 500 verschiedenen, von ddm1 abgeleiteten epigenetischen rekombinanten Inzuchtlinien (epiRILs) kreuzten. Wir führten Methylom-, Transkriptom- und phänotypische Profilerstellung für eine Reihe von Eltern-Hybrid-Trios durch, und unsere Daten zeigten, dass die Methylome der Hybriden einem umfassenden Umbau unterzogen wurden. Diese Umgestaltungsereignisse im Methylom der Hybriden waren nicht isoliert, sondern traten in weit voneinander entfernten Regionen auf. Obwohl die genauen Regulierungsmechanismen noch nicht vollständig geklärt sind, deuten unsere Daten darauf hin, dass trans-wirkende 24nt small RNA (sRNA) eine Schlüsselrolle in diesem Prozess spielt. Wichtig ist, dass wir gezeigt haben, dass die stabile Segregation elterlicher perizentromerischer DMRs in den Hybriden nicht nur lokal, sondern auch in distalen Regionen Methylierungsänderungen bewirkten, die Tausende von Zielstellen im gesamten Genom betrafen. Bemerkenswerterweise waren viele dieser trans-induzierten Veränderungen mit der Expression nahe gelegener Gene verbunden und standen in signifikantem Zusammenhang mit phänotypischer Heterosis. mit phänotypischer Heterosis verbunden.

Basierend auf unseren Forschungsergebnissen legen wir ein überzeugendes Modell nahe, bei dem elterliche Unterschiede in den DNA-Methylierungsmustern, insbesondere in den heterochromatinreichen perizentromerischen Regionen, eine wichtige Rolle bei der Reorganisation der Methylome und Transkriptome von Hybriden spielen. Diese Veränderungen tragen zu nicht additiven phänotypischen Effekten bei, die das Phänomen der Heterosis unterstreichen. Diese Effekte treten in isogenem Pflanzenmaterial auf und unterstreichen die Bedeutung epigenetischer Variation als molekulare Determinante dieses Phänomens.

List of Figures

1.1 A transcription fork model for RNA-directed DNA methylation 3
1.2 Classification of DNA methylation remodeling events in F1 hybrids 6
1.3 Trans-chromosomal methylation and demethylation models for DMR events . . . 6
1.4 Construction of the Col-wt epiRILs 12
1.5 Crossing scheme 14
1.6 An epigenetic-QTL approach 15

2.1 Selection of epiRILs 18
2.2 Images of the carriers inside the growth facility on DAS 18 19
2.3 Normalized leaf area measurements 20

3.1 Schematic depiction of the crosses 28
3.2 SNP density for the msCol maternal line 29
3.3 Frequency of DMRs between msCol and a Col-wt 30
3.4 Global methylation levels for the epiRILs 31
3.5 Schematic model that was used to characterize methylome remodeling regions . . 31
3.6 Representative browser shots for family 188 33
3.7 Statistics for the methylome remodeling categories 34
3.8 Mechanistic insights of the remodeling regions 35
3.9 DMR frequency distribution and length distribution of sRNA reads 36
3.10 Distribution of sRNAs known targets length 36
3.11 Frequency distribution of line being a NAD 38
3.12 Pairwise-correlation analysis of remodeling regions 39
3.13 Composition of significant NAD correlations 40
3.14 Correlation structure of the distally located regions 41
3.15 A haplotype map of the 169 epiRIL lines 42
3.16 Significant QTL^{epi} regions 43
3.17 Genome wide associations between DMR markers and NAD regions 45
3.18 DMRs trigger transcriptional and phenotypic heterosis 46
3.19 epiHybrid methylation divergence correlates with expression change 47
3.20 Description of causal models 48
3.21 Distances between NAD-QTLs targets in distal regions and their corresponding confidence interval boundaries 48

3.22 Epigenome-wide association study 50
3.23 Genome-wide QTL scan for LA 50
3.24 Genetic variation does not affect methylome remodeling in the epiHybrids 51

4.1 A TCM-proximity model for SMR-gain events 57
4.2 A “distal” model for SMR-loss events 58

List of Tables

1.1 Examples of studies focusing on epigenetic remodeling mechanisms in diploid plants (Kakoulidou and Johannes 2023). 5

2.1 Example of alignment and coverage summarising statistics 21
2.2 Remodeling senario at each 200bp region 23

3.1 Genes where methylation divergence correlates with expression divergence. . . . 52
3.2 Genes where the NAD-QTL epigenotypes correlates with expression divergence. 53

Chapter 1

Introduction

Heterosis describes the superior phenotype of an F1 offspring compared to the phenotype of its two inbred parents (Birchler et al. 2010; Schnable and Springer 2013), most notably for traits such as biomass, seed yield, or developmental rate (Chen 2010). Due to its potential for enhancing crop production (Schnable and Springer 2013), heterosis, which was first described by Shull in 1908, has been extensively investigated globally both in the academic and in the commercial sector, with a focus on its molecular and phenotypic characteristics. Yet, despite its importance, the underlying molecular mechanisms remains so far poorly understood (Schnable and Springer 2013).

Various genetic hypotheses have been proposed to explain heterosis, all of which revolve around classical models of dominance, over-dominance, or epistasis. The dominance model posits that the enhanced performance of hybrids results from the presence of slightly deleterious recessive alleles at multiple loci in the inbred parents, which are complemented by dominant alleles in the hybrid (Jones 1917; Schnable and Springer 2013). According to the over-dominance model, the superior performance of hybrids is attributable to heterozygosity at individual loci (Crow 1948). All of these models suggest that heterosis is expected to increase with the genetic distance between the parental lines. This prediction is backed by empirical evidence, which indicates that heterosis is generally more significant in crosses between different species than those within the same species (East 1936; Chen 2010).

Despite the longstanding predictions of these models, numerous exceptions have been identified over time, challenging their generality and paving the way for alternative explanations. One such intriguing model suggests that epigenetic variations, rather than genetic differences, may be responsible for triggering heterosis (Chen 2010; Lauss et al. 2018; Dapp et al. 2015; Groszmann et al. 2013). This hypothesis is supported by evidence that F1 hybrids undergo significant DNA methylation remodeling compared to their parental lines (Kakoulidou et al. 2024; Greaves et al. 2012a; Zhang et al. 2016) and that experimental manipulation of DNA methylation pathways can affect the heterotic potential of hybrids (Kawanabe et al. 2016; Lauss et al. 2018; Kakoulidou et al. 2024; Rigal et al. 2016).

The remodeling of DNA methylation patterns occurs alongside other chromatin modifications, which collectively alter the regulatory environment of hybrid genomes and potentially provide a molecular foundation for heterosis (Kakoulidou and Johannes 2023). Notably, similar epigenetic alterations have been documented in the F1 progeny of isogenic parents that were genetically modified to have different DNA methylation patterns (Dapp et al. 2015; Lauss et al. 2018), underlining that these events are not just consequences of genetic variation, and thus call for independent mechanistic explanations.

Several studies have examined the DNA methylation patterns in F1 hybrids and their parents in detail. These studies have raised several questions, such as what triggers DNA methylation remodeling, how do these changes lead to regulatory and phenotypic alterations, and how can epigenetic information be utilized as a biomarker for predicting heterosis. However, there are still gaps in our understanding of the molecular mechanisms involved in DNA methylation remodeling in F1 hybrids (Kakoulidou and Johannes 2023).

1.1 DNA methylation

DNA cytosine methylation is an epigenetic modification found in eukaryotes that involves adding a methyl group (CH3) to a cytosine nucleotide. This modification occurs in CG, CHG, and CHH sequence contexts, and is commonly found in pericentromeric heterochromatic regions of chromosomes, where it is associated with the repression of transposable elements (TEs) and repetitive sequences. The RNA-directed DNA methylation (RdDM) pathway, which involves 24-nucleotide small RNAs (sRNAs) acting as guide molecules for DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), primarily catalyzes *de novo* methylation in all three sequence contexts (Figure 1.1). Pol IV and Pol V RNA polymerases are also involved in the production of sRNAs and their targeting to specific loci (Matzke et al. 2009; Law and Jacobsen 2010). In addition to their role in heterochromatin, sRNAs also play a role in gene expression regulation through post-transcriptional silencing mechanisms, as well as in plant development, reproduction, and phenotypic plasticity (Borges and Martienssen 2015).

DNA methylation is maintained by context-specific pathways once it's established. CG context sites are recognized by the VIM1 family of proteins, which recruits METHYLTRANSFERASE 1 (MET1) to copy the template and catalyze CG methylation on the newly synthesized strand. Loss of MET1 results in complete genome-wide loss of CG methylation (Saze et al. 2003). CHG methylation is mostly maintained by the plant-specific methyltransferase CMT3 (Law and Jacobsen 2010), which acts in a self-reinforcing loop with histone H3 lysine 9 demethylation (H3K9me2) and the histone methyltransferase SUPPRESSOR OF VARIEGATION 3-9 HOMOLOGUE 4 (SUVH4) (Johnson et al. 2007). At a subset of CHG and CHH sites, methylation is also maintained by the *de novo* activity of CMT2, which requires H3K9me2 (Law and Jacobsen 2010). Several DNA methylation mutants have revealed significant cross-talk between these pathways (To and Kakutani 2022).

In addition to RdDM, a key factor in maintaining DNA methylation patterns is the chromatin remodeler DECREASE in DNA METHYLATION 1 (DDM1). DDM1 regulates methylation by providing access for DNA methyltransferases to H1-containing histones, primarily within long TEs of heterochromatic regions (Zemach et al. 2013). Loss of DDM1 results in a 70% reduction in overall DNA methylation and a widespread over-accumulation of TE-related transcripts (Kakutani et al. 1995; Vongs et al. 1993; Lippman et al. 2004). Interestingly, DDM1 is also required to

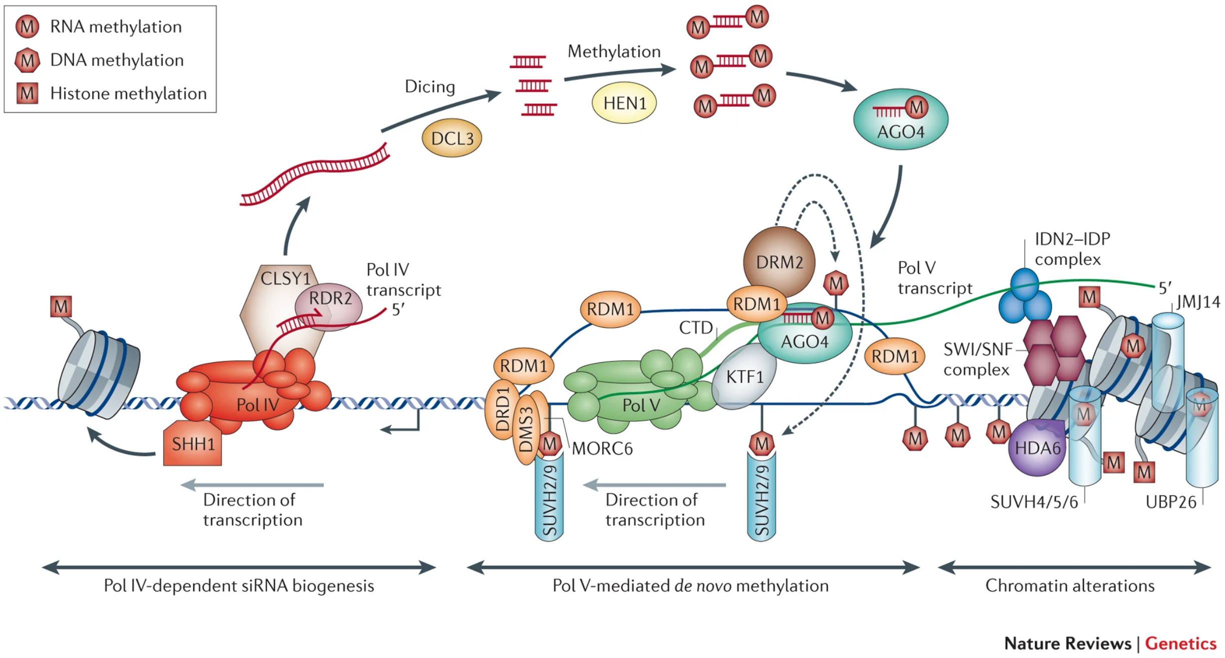

RNA methylation
DNA methylation
Histone methylation
Dicing
DCL3
Methylation
HEN1
AGO4
M
CLSY1
RDR2
Pol IV transcript
5′
Pol IV
SHH1
Direction of transcription
DRM2
RDM1
CTD
AGO4
KTF1
RDM1
Pol V
DRD1
DMS3
MORC6
SUVH2/9
Direction of transcription
SUVH2/9
Pol V transcript
RDM1
IDN2–IDP complex
5′
JMJ14
SWI/SNF complex
HDA6
SUVH4/5/6
UBP26
Pol IV-dependent siRNA biogenesis
Pol V-mediated *de novo* methylation
Chromatin alterations
Nature Reviews | Genetics

Figure 1.1: a) A transcription fork model for RNA-directed DNA methylation (RdDM) (Matzke and Mosher 2014).

maintain histone H3 methylation patterns, as loss of DDM1 in the mutant is accompanied by the replacement of lysine 9 methylation with lysine 4 methylation (Soppe et al. 2000; Gendrel et al. 2002). At the phenotypic level, altered flower morphology, late flowering, and low sterility are among the morphological phenotypes of the *ddm1* mutant (Kakutani et al. 1996). In addition, impermanent loss of DDM1 can induce heritable epialleles that segregate independently and contribute to phenotypic heritability (Kakoulidou and Johannes 2023).

1.2 Classification of DNA methylation remodeling events

A number of experimental studies have examined epigenetic changes in F1 hybrids, including in *A. thaliana* (Arabidopsis), *Oryza sativa* (rice), *Brassica napus* (oilseed rape), *Zea mays* (maize) and *Cajanus cajan* (pigeonpea) (Kakoulidou and Johannes 2023) (Table 1.1). These studies focus mainly in the DNA methylation remodeling but also included sRNA and RNA expression data sources, as well as phenotypic data in a way to delineate mechanistic causes and functional consequences (Kakoulidou and Johannes 2023).

The majority of studies have used Arabidopsis as a model organism. Although this primarily selfing species may not be representative of obligate outcrossers, its small genome size and extensive resources for (epi)genetic research enable detailed mechanistic investigations. Furthermore, several fundamental epigenetic observations in Arabidopsis hybrids have been found to be applicable to more complex outcrossing species like maize (Barber et al. 2012), indicating that Arabidopsis is a justifiable model.

The initial use of Arabidopsis was motivated by early observations that genetically similar ecotypes (e.g., C24 X Ler) can exhibit extensive vegetative heterosis for traits as rosette diameter and biomass. As a result, it has been hypothesized that heterotic phenotypes are triggered by mostly epigenetic, rather than genetic, differences between the parents. More recent attempts to examine the role of parental epigenetic differences in facilitating DNA methylation remodeling and heterosis in F1 hybrids have been undertaken by Dapp et al. (2015); Rigal et al. (2016); Lauss et al. (2018); **?**. These studies examined F1 hybrids from crosses between isogenic parents engineered to differ only in their epigenetic profiles. Their parental lines were selected from existing panels of *ddm1* or *met1*- derived epigenetic recombinant inbred lines (epiRILs; (Johannes et al. 2009; Reinders et al. 2009)).

The use of isogenic parental lines demonstrates the value of this model for investigating the role of epigenetic differences in heterosis, with potential implications for more complex outcrossing species such as maize (Barber et al. 2012). Regardless of the organism, previous studies have found that most parental methylation states are inherited additively in hybrids, including those conducted by Zhang et al. (2016); Lauss et al. (2018); Ma et al. (2021). These findings suggest that the methylation states of both parental alleles are maintained stably and independently, as depicted in Figure 1.2.

However, it's important to note that a significant proportion of hybrid genomes undergo changes in their DNA methylation levels relative to their parental lines (Zhang et al. 2016; Shen et al. 2012; Greaves et al. 2012a; He et al. 2010; Greaves et al. 2012b). These changes are known as methylation remodeling events, which can occur as either methylation gains or losses in regions of the genome where the two parents have differential methylation patterns (known as differentially methylated regions or DMRs) (Shen et al. 2012; Greaves et al. 2012b; Chodavarapu et al. 2012) or similar methylation patterns (known as similarly methylated regions or SMRs) (Figure 1.2; (Zhang et al. 2016; Lauss et al. 2018; Ma et al. 2021; Li et al. 2018).

Methodologically, categorizing the various types of DNA methylation remodeling events in hybrids is not a simple task. To distinguish between monoallelic and biallelic methylation changes in hybrids, allele-specific DNA methylation data is necessary. This data can be acquired for genomic regions where the two parents are genetically different, as the parental origin of the sequencing reads can be identified (Kakoulidou and Johannes 2023). However, in non-polymorphic regions of genetically distinct parental lines or isogenic F1 hybrid systems (e.g., *ddm1*-derived or *met1*-derived epiHybrids), such information is unavailable. In these cases, classification must rely on analyzing the changes in DNA methylation levels in the hybrids relative to the parents (Kakoulidou and Johannes 2023).

Although not precise, this approach has proven effective, especially for parental regions with significant methylation levels. Indeed, categorizing remodeling events based on changes in methylation levels yields similar frequencies of the different types of remodeling events as allele-specific analyses. The above classification encompasses the primary types of remodeling events seen in F1 hybrids and has proven helpful in developing mechanistic models (Kakoulidou and Johannes 2023). In reality, classification is much fuzzier, and should be understood as existing on a continuum of remodeling possibilities (e.g., DMR gain can occur on both alleles). Nonetheless, this classification has been useful for guiding the development of mechanistic models (Kakoulidou and Johannes 2023).

Table 1.1: Examples of studies focusing on epigenetic remodeling mechanisms in diploid plants (Kakoulidou and Johannes 2023).

Reference	Species	Epigenetic mechanism
Zhang et al. (2016)	Arabidopsis	DNA methylation & sRNAs
Shen et al. (2012)	Arabidopsis	DNA methylation & sRNAs
Greaves et al. (2012b)	Arabidopsis	DNA methylation & sRNAs
Chodavarapu et al. (2012)	Rice	DNA methylation & sRNAs
Shen et al. (2017)	Oilseed rape	DNA methylation & sRNAs
Groszmann et al. (2011b)	Arabidopsis	DNA methylation & sRNAs
Li et al. (2012)	Arabidopsis	sRNAs
Sinha et al. (2020)	Pigeonpea	DNA methylation & sRNAs
Barber et al. (2012)	Maize	sRNAs
Lauss et al. (2018)	Arabidopsis	DNA methylation
Kakoulidou et al. (2024)	Arabidopsis	DNA methylation & sRNAs
Rigal et al. (2016)	Arabidopsis	DNA methylation & sRNAs
Dapp et al. (2015)	Arabidopsis	Epigenetic regulation of transcription
Kawanabe et al. (2016)	Arabidopsis	DNA methylation & sRNAs
Greaves et al. (2016)	Arabidopsis	DNA methylation & sRNAs
Ma et al. (2021)	Rice	DNA methylation & sRNAs
Greaves et al. (2014)	Arabidopsis	DNA methylation
He et al. (2010)	Rice	DNA methylation & sRNAs histone modification

1.3 Mechanisms of DNA methylation remodeling

It is clear that the majority of DNA methylation remodeling events in F1 hybrids seem to occur in parental SMRs rather than in DMRs (Figure 1.2c) (Zhang et al. 2016; Greaves et al. 2012a,b; Lauss et al. 2018). However, only a small fraction of parental genomes are typically differentially methylated, thus making the number of DMR-based remodeling events present a significant enrichment. That means that parental DMRs are more likely to have methylation changes events in the hybrids and are better studied so far (Kakoulidou and Johannes 2023).

1.3.1 DMR-gain events

DMR-gain events have been termed as ”trans-chromosomal methylation (TCM)” (Greaves et al. 2012b). This term describes a molecular model where two parental alleles interact in the hybrids via 24nt sRNA. sRNA initially produced in the methylated parental allele, target subsequently

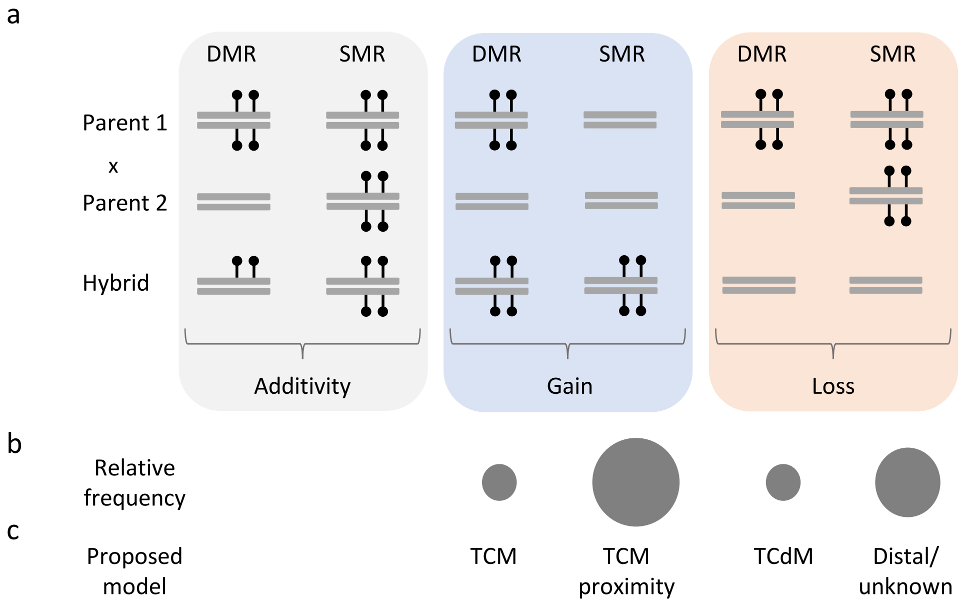

a
DMR
SMR
DMR
SMR
DMR
SMR
Parent 1
x
Parent 2
Hybrid
Additivity
Gain
Loss
b
Relative frequency
c
Proposed model
TCM
TCM proximity
TCdM
Distal/ unknown

Figure 1.2: a) Schematic representation of the different remodeling events. b) Relative frequency of remodeling events across the genome approximated from (Kakoulidou et al. 2024; Lauss et al. 2018; Zhang et al. 2016). Larger circles denote higher frequency. c) Proposed mechanistic models for remodeling events (Kakoulidou and Johannes 2023).

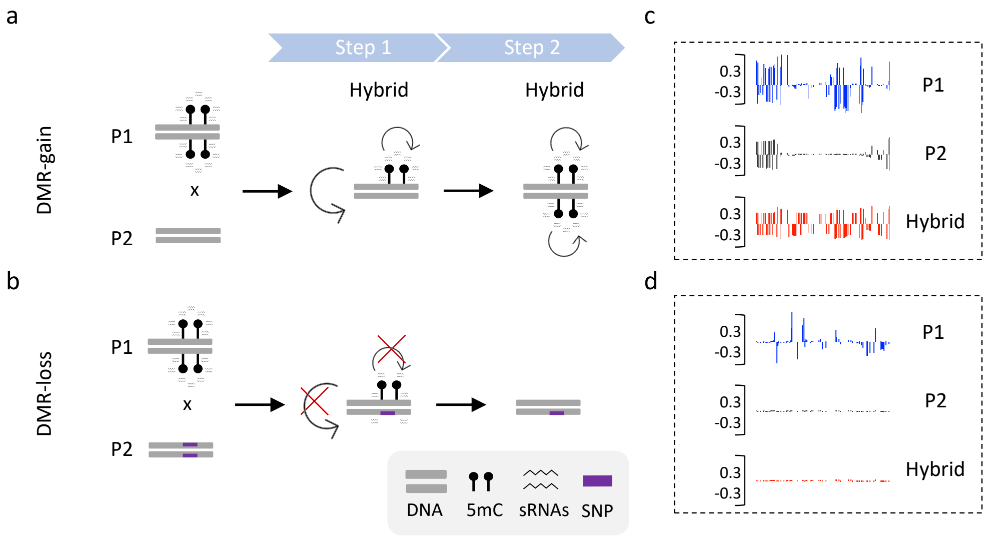

a
Step 1
Step 2
Hybrid
Hybrid
DMR-gain
P1
x
P2
b
DMR-loss
P1
x
P2
DNA
5mC
sRNAs
SNP
c
0.3
-0.3
P1
P2
Hybrid
d
P1
P2
Hybrid

Figure 1.3: Trans-chromosomal methylation (a) and demethylation (b) models for DMR events (Kakoulidou and Johannes 2023).

the unmethylated allele for *de novo* methylation (Figure 1.3). Once established, the methylation status of the recipient allele is then maintained independently. Here the term ”trans” emphasize the allelic interaction between two homologous chromosomes and it shouldn’t been confused with the term ”trans” in genetics that described interactions in two distal regions in the genome.

Some of the most important observations supporting this model involve 24nt sRNAs. Firstly, TCM loci are targets of 24nt sRNA and are often accompanied by differential sRNA abundance (Greaves et al. 2016). Next, the monoallelic methylation gains in the hybrids tend to co-occur with a gain of 24nt sRNA on the same allele (Zhang et al. 2016). Lastly, these remodeling events require the presence of 24nt sRNA, as they are largely abolished in RdDM mutants (Greaves et al. 2016; Zhang et al. 2016).

It is noteworthy that the development of TCM occurs gradually during the growth of plants, as stated by Greaves et al. (2014). This may suggest that the strengthening of methylation on the recipient allele is partially reliant on replication. However, the timing of TCM regions varies within the same system, likely due to differences in the intensity of the TCM signal that involves sRNAs, as also noted by Greaves et al. (2014).

Due to the close association between DNA methylation and various histone modifications, alterations in other epigenetic marks frequently accompany TCM events. For instance, as demonstrated by Greaves et al. (2014), specific gains in methylation induced by TCM are accompanied by a reduction in H3K9ac (Groszmann et al. 2013). This implies that TCM can trigger changes in higher-level chromatin states in hybrids. However, in other cases, the opposite is true. The FLOWERING WAGENINGEN A (FWA) gene, a well-known regulator of flowering time in plants, provides an example. Methylation of repeat elements in the promoter and the 5’ untranslated region of FWA silences its transcription, which is mediated by sRNA (Kinoshita et al. 2007). While sRNA sites can be targeted for TCM, this is not the case when the recipient allele’s FWA is marked by active chromatin and expressed (Chan et al. 2004). This suggests that active chromatin hinders sRNA-directed TCM (Groszmann et al. 2011a).

The TCM model presented above is based heavily on prior research on paramutation in maize (Chandler 2007). In fact, the initial stage of paramutation involves a TCM process, as the methylated inducing allele functions as a template for Pol IV and sRNA production, allowing its epigenetic state to be imposed on the sensitive (unmethylated) allele. sRNAs produced by the inducing allele reinforce its silenced state in *cis* and subsequently direct *de novo* methylation of the other allele (Hövel et al. 2015).

As in paramutation, DMR-gain events can be inherited beyond the F1 generation, making it a source of heritable epialleles (Greaves et al. 2014; Zhang et al. 2016). However, this inheritance pattern can be disrupted by perturbations of the RdDM pathways (Zhang et al. 2016). Paramutable alleles that are often targeted by other methylation pathways like MET1 or DDM1 seem to be more stable and are also enriched for CG context, indicating the transgenerational stability (Greaves et al. 2016).

In contrast, the instability of numerous TCM-induced methylation states is an intriguing phenomenon in its own regard, as it could potentially underlie the breakdown of hybrid vigor commonly observed in the F2 or F3 generation (Greaves et al. 2014). While this hypothesis is plausible, it necessitates demonstrating a direct connection between TCM and phenotypic heterosis, as well as ruling out alternative explanations such as the loss of epistatic genotype combinations due to segregation or recombination, both of which are challenging to test experimentally (Kakoulidou and Johannes 2023).

1.3.2 DMR-loss events

The opposite of TCM has also been proposed as an explanation for DMR loss events, referred to as trans-chromosomal demethylation (TCdM) by analogy (Greaves et al. 2012b). This theory is supported by the observation that a reduction in matching 24-nt sRNA on the same allele accompanies monoallelic methylation losses (Zhang et al. 2016) (Figure 4.1). However, this is only correlational evidence, and the molecular basis of the ”trans-chromosomal” signal is not well understood. Additionally, TCM and TCdM events are not likely to be separate processes, as TCdM events are more frequent in regions with higher levels of CHH methylation, suggesting a different mechanism than the one responsible for the establishment of TCM.

One potential explanation for DMR loss events was proposed by Zhang et al. (2016) based on their observation that these events tend to occur in regions with high genetic variation among the parental alleles. The authors suggested that the lack of sequence homology between the methylated and unmethylated alleles hinders efficient targeting of sRNA, leading to a dilution of sRNA copies and weakened methylation reinforcement on the methylated allele. While this mechanism may hold for highly polymorphic sites, it is not clear how this could be possible in isogenic lines and this raises the possibility that the trigger can originate from remodeling events in distal regions.

1.3.3 SMR-gain and loss events

That remodeling events can occur at positions in the genome where parental regions differ and allows a cross-talk between parental alleles is straightforward. However, in some studies methylation gain or loss also occurs in regions where parental alleles are similarly methylated (Lauss et al. 2018; Zhang et al. 2016; Ma et al. 2021). Even though it is not clear how these remodeling events are triggered, the signal must come from a distal region outside of the homologous alleles (Kakoulidou and Johannes 2023). Nevertheless, this is an area of research that is largely unexplored.

1.4 Links to non-additive gene expression and heterosis

Studies have consistently shown that DNA methylation remodeling in F1 hybrids is associated with non-additive changes in gene expression (Chen et al. 2022; Greaves et al. 2012a; He et al. 2010; Sinha et al. 2020; Groszmann et al. 2011a), particularly in promoter regions. Gain or loss of methylation in proximity to genes often correlates with up- or down-regulation of gene expression, respectively, and this association appears to depend on specific developmental stages and tissue types (Fujimoto et al. 2018). The affected genes are often enriched in pathways regulating circadian rhythm, hormones, and metabolism (Ni et al. 2009), hormones (Chen et al.

2022; Shen et al. 2012; Sinha et al. 2020), which is consistent with the observed changes in photosynthetic activity, growth, and flowering time that often occur in F1 hybrids.

Methodologically, it remains challenging to establish a link between DNA methylation remodeling events and phenotypic heterosis and to determine whether DNA methylation drives gene expression or vice versa. However, research with DNA methylation mutants has shown that experimental manipulation of DNA methylation pathways can affect the heterotic potential of F1 hybrids for key traits such as vegetative biomass, pathogen resistance, and flowering time (Johannes et al. 2009; Kawanabe et al. 2016; Reinders et al. 2009).

For example, loss of DDM1 results in a significant reduction in heterosis for rosette diameter (Kawanabe et al. 2016). However, the role of RdDM in phenotypic heterosis is less clear, which is surprising given its central role in TCM. Studies by Barber et al. (2012); Zhang et al. (2016); Kawanabe et al. (2016) report that loss of Pol IV, Pol V, or MODIFIER OF PARAMUTATION 1 (MOP1) does not significantly impact F1 heterosis. In contrast, Shen et al. (2012)) found significant effects when abolishing sRNA production by knocking down the Arabidopsis RNA methyltransferase HUA ENHANCER1 (HEN1). Although technical differences may contribute to these discrepancies, they suggest the possibility that phenotypic heterosis is at least partially independent of RdDM.

1.5 Outstanding questions in the field

Hybrids exhibit significant alterations in DNA methylation and sRNA patterns compared to their parental lines. The majority of DNA methylation remodeling in F1 hybrids is known to occur in parental SMRs, rather than DMRs. However, the underlying mechanisms for these observations are still not entirely clear and may involve distally-acting factors, among others. Causal loci that facilitate these distal effects remain unidentified, and it is currently unclear how genome-wide epigenetic remodeling in hybrids contributes to the emergence of heterosis at the phenotypic level. A key challenge in studying the relationship between parental and hybrid epigenomes is to isolate it from genetic variation. This is necessary to attribute any functional and phenotypic changes in hybrids to the epigenetic state of parents, rather than to DNA sequence variations (Kakoulidou and Johannes 2023). Indeed research has revealed that heritable morphological variations in Arabidopsis plants can arise exclusively due to epigenetic factors, as evidenced by studying isogenic epigenetic recombinant inbred lines (epiRILs). To investigate the impact of parental epigenetic variation on F1 heterosis, the same epiRILs have been used to create F1 epigenetic hybrids, which are referred to as epiHybrids hereafter.

1.6 Epigenetic Recombinant Inbred Lines (epiRILs)

Although piRILs are nearly isogenic, they exhibit mosaicism in their epigenomes. Two populations of Arabidopsis epiRILs have been established by crossing wild-type Columbia accession (Col-wt) with wild-type lines that bear a mutation in either MET1-3 or DDM1-2 (Johannes et al. 2009; Reinders et al. 2009).

1.6.1 The *met1-3* epiRILs

The *met1-3* derived population was initiated through a cross between a wild-type plant and a plant carrying a loss of function mutation in MET1, both of which were Col-wt ecotypes, as reported by Reinders et al. (2009). As MET1 is crucial for maintaining CG methylation patterns during DNA replication, the *met1-3* null mutant is characterized by a near-complete loss of CG methylation (Saze et al. 2003). In this experiment, a *met1-3* plant was crossed with a wild-type plant, and segregating MET1 homozygous F2 plants were selected and propagated through self-pollination and single-seed descent. The inbreeding of 96 MET1 F2 lines led to the creation of the ”epiRILs.” Since the parental DNA methylation differences are conserved, the resulting epiRIL population displays a mosaic of DNA methylation patterns. According to the crossing experiment, the parental epialleles that were not selected, are inherited in a Mendelian fashion and are segregated with a 1:1 ratio (Reinders et al. 2009).

At the phenotypic level, the *met1-3* mutant has demonstrated developmental abnormalities such as short plant stature, late flowering, altered flower morphology, and reduced fertility, which are generally more frequent and more severe with further inbreeding (Mathieu et al. 2007). Interestingly, some of these traits seemed to be inherited from the *met1-3* derived epiRILs providing the first insight into the extent of phenotypic divergence produced by randomly combining epialleles (Reinders et al. 2009).

1.6.2 The *ddm1-2* epiRILs

The other epiRIL population was created using two closely related parents of the same accession (Columbia, Col), similar to the *met1-3* derived epiRILs. However, in this case, a DNA hypomethylation mutant with a mutation at the DDM1 locus was used. Since DDM1 is primarily involved in maintaining DNA methylation and silencing repeat elements (Kakutani et al. 1995; Vongs et al. 1993; Lippman et al. 2004), *ddm1-2* mutants exhibit a 70% reduction in DNA methylation overall in all three contexts of heterochromatin transposable elements (TEs) (Kakutani et al. 1995; Vongs et al. 1993; Lippman et al. 2004).

To conduct this experiment, a Col-wt plant homozygous for the *ddm1-2* mutant allele was used as the paternal parent, and a Col-wt plant homozygous for the wild-type DDM1 allele was used as the maternal parent. From the backcross progeny, 500 individual DDM1/DDM1 plants were selected, and after six rounds of inbreeding propagation by single seed descent, the *ddm1-2*-derived epiRIL population was created (Johannes et al. 2009) (Figure 1.4). Although each epiRIL line has a very similar genome, they differ significantly in their DNA methylation levels due to the stable inheritance of DNA methylation variants induced by *ddm1-2* in a 1:3 ratio in each epiRIL line (Johannes et al. 2009).

Previous sequencing of the epiRILs has revealed increased mobilization rates of some transposable element families (CACTA, ATCOPIA93, ATENSPM3, and VANDAL21). However, the majority of these are specific to each epiRIL family and have no significant phenotypic effect (Quadrana et al. 2019; Cortijo et al. 2014; Johannes et al. 2009; Marí-Ordóñez et al. 2013).

Furthermore, there is currently no evidence to suggest that the epiRILs segregate any other types of genetic variants originating from the initial *ddm1-2* founder (Lauss et al. 2018), except for a well-described 2Mbp inversion on chromosome 2 (Zhang et al., under review). This inversion was also detected in the genomes of the *ddm1-2*-derived epiRILs (Zhang et al., under review) and in the genomes of *ddm1-1* (Vongs et al. 1993), but was not detected in another sibling *ddm1-1* mutant more recently (He et al. 2021), indicating that it can be distinguished from the DDM1 mutated locus via genetic segregation. Additionally, two other medium-sized duplications were recently detected in *ddm1* siblings (Zhang et al., under review and Yi and Richards (2008), but these appear to be specific to certain siblings.

In contrast to the *met1-3* mutant, which showed no significant developmental defects, the *ddm1-2* mutant used for the crosses displayed minor developmental defects such as small leaf size, increased apical dominance, and late flowering. However, these phenotypes became progressively more severe in advancing generations, indicating a potential negative impact on the genetic fitness of the mutant (Kakutani et al. 1996). Nevertheless, despite the variation in the severity of the phenotypes, the epiRILs exhibited high heritability variation for a wide range of important agricultural traits. These traits included plant height, root length, flowering time, rosette size, resistance to pathogens, and phenotypic plasticity under salt stress (Johannes et al. 2009; Cortijo et al. 2014; Roux et al. 2011; Furci et al. 2019; Kooke et al. 2015).

1.6.3 Epigenetic Hybrids (epiHybrids)

To address the challenge of disentangling genomic and epigenomic sources of variation in hybrid plants, Lauss et al. (2018) used a unique approach. They generated Arabidopsis F1 hybrids with epigenetic differences (epiHybrids) by crossing near-isogenic parental lines that were divergent in their epigenetic profiles. Specifically, they used Col-wt as the maternal parent and the near-isogenic ddm1-2-derived epiRILs (Johannes et al. 2009) as the paternal parent (Figure 1.4). Thus, each epiHybrid inherited one wildtype chromosome from the recurrent female parent and one epiRIL chromosome from their respective paternal parents. Since the parental lines had virtually the same DNA sequences, the epiHybrids differed only on their paternal methylome contributions.

Extensive phenotypic screens for phenotypic traits such as seed yield, leaf area, plant height, and flowering time revealed widespread heterosis in the epiHybrids. This finding provided strong evidence that methylation differences between parental lines can trigger heterosis in hybrid offspring, independently of DNA sequence determinants (Lauss et al. 2018).

Of particular significance, the researchers identified nonadditively methylated regions throughout the genome in four epiHybrids. These regions were more frequently observed in areas where a parental line of the epiRIL carried a hypomethylated *ddm1*-derived epihaplotype, in comparison to windows where both parents had the Col-wt haplotype. About 2.1% of regions across the four epiHybrids exhibited a negative deviation from the MPV, while approximately 2.4% showed a positive deviation. In cases where the parents exhibited differential methylation, lower methylation levels were observed, whereas regions with similar methylation in the parents typically displayed increased methylation levels. Taken together, these observations suggest that *de novo* DNA methylation may be occurring due to sRNAs derived from nonallelic homologous sequences, primarily in regions where parental lines have similar methylation patterns.

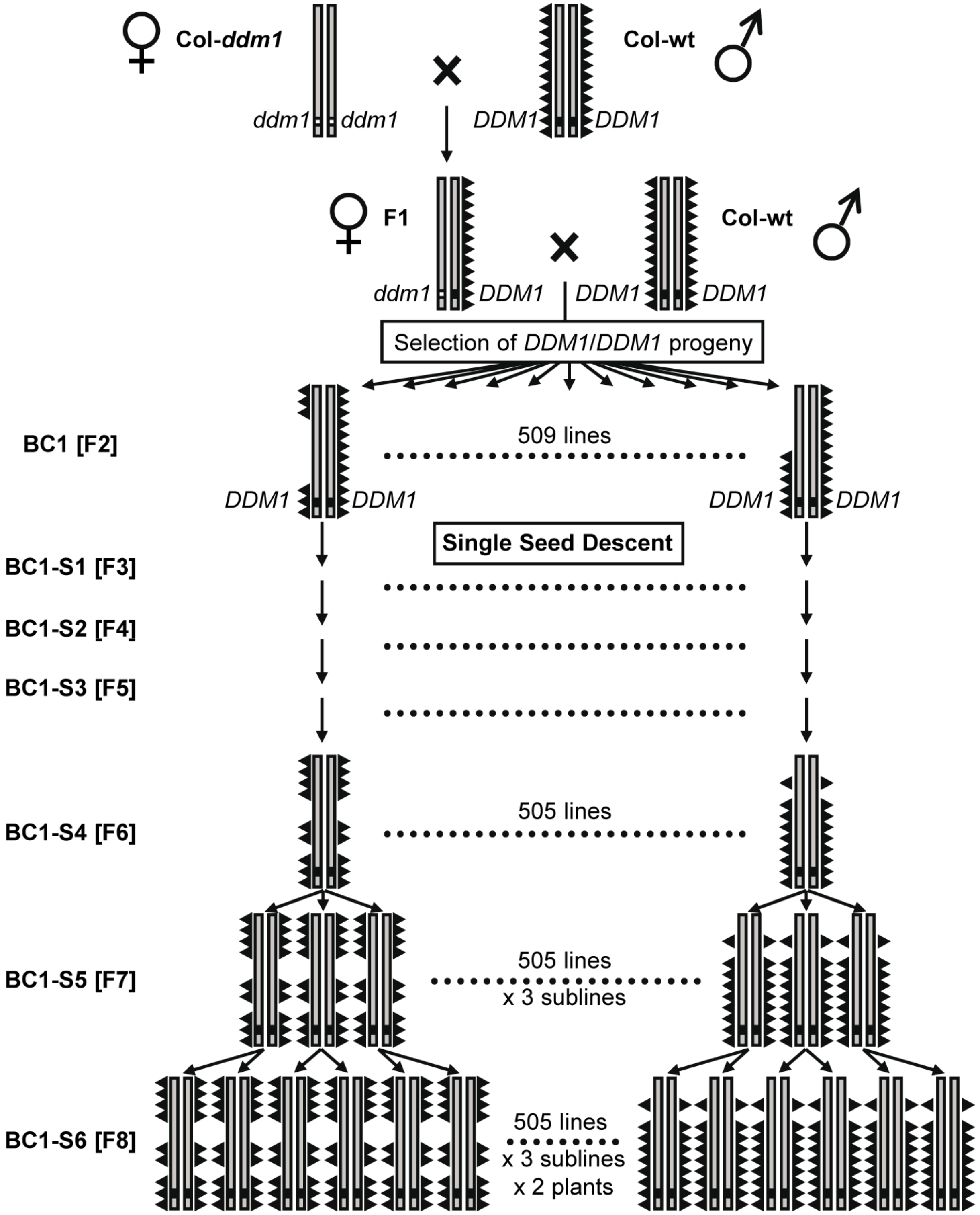

♀ Col-*ddm1*
×
Col-wt ♂
ddm1 *ddm1*
DDM1 *DDM1*
♀ F1
×
Col-wt ♂
ddm1 *DDM1*
DDM1 *DDM1*
Selection of *DDM1*/*DDM1* progeny
BC1 [F2]
509 lines
DDM1 *DDM1*
DDM1 *DDM1*
Single Seed Descent
BC1-S1 [F3]
BC1-S2 [F4]
BC1-S3 [F5]
BC1-S4 [F6]
505 lines
BC1-S5 [F7]
505 lines
x 3 sublines
BC1-S6 [F8]
505 lines
x 3 sublines
x 2 plants

Figure 1.4: Construction of the Col-wt epiRILs. Grey bars represent the *A. thaliana* genome and triangles represent DNA methylation. Except at the DDM1 locus located on chromosome 5, the two parents are near isogenic; they differ however in their levels of DNA methylation (Johannes et al. 2009).

These findings shed new light on the mechanisms underlying DNA methylation patterns and their effects on hybrid offspring. They also suggest that hypomethylated haplotypes, such as those observed in the *ddm1*-derived epihaplotype, may contribute to heterosis in epiHybrids. Around the same time, Reinders et al. (2009) drew similar conclusions at the phenotypic level using the met-derived epiRILs. Hybrids between the epiRILs and Col-wt displayed hybrid vigor for rosette size, demonstrating that divergent epigenome combinations can trigger heterosis. They also observed a parent-of-origin effect, with heterotic phenotypes appearing only when the epiRIL was the maternal parent and transmitted its hypomethylated epigenome.

To further explore the underlying genetic mechanisms of heterosis, Reinders et al. (2009) conducted genome-wide transcription profiling of the epiHybrids. Their analysis revealed a number of genes with altered regulation that may contribute to the observed heterosis. These results provided insight into the complex interplay between epigenetic modifications and gene expression in determining hybrid vigor. Overall, their work highlighted the potential of using epiRILs as a model system for studying the role of epigenetics in hybrid performance and may have important implications for improving crop yields through the manipulation of epigenetic marks.

1.7 Epigenetic diversity as a predictive marker of hybrid's performance

Recent studies have attempted to establish a connection between phenotypic heterosis and specific characteristics of parental DNA methylomes. Lauss et al. (2018) demonstrated a proof of concept by using an epigenetic quantitative trait locus mapping technique QTL^{epi} in a large collection of the *ddm1*-derived epiHybrids.

DMRs have been identified between the epiRIL founder lines, the *ddm1-2* and Col-wt, that are stably inherited through at least 10 rounds of meiosis in the paternal epiRIL lines (Colomé-Tatché et al. 2012). These highly stable DMRs serve as physical markers that have been utilized to create a recombination map (Colomé-Tatché et al. 2012; Cortijo et al. 2014). At each DMR marker, any given epiRIL parent may be epihomozygous for either the wild-type methylated state (MM) or the *ddm1-2*-like state (UU). Interestingly, the genetic map resulting from these markers is comparable in length to those derived from classical Arabidopsis crosses, indicating that the hypomethylated loci that segregate in the *ddm1* epiRILs do not appear to have an impact on global meiotic recombination rates (Catoni and Cortijo 2018).

In a study conducted by Lauss et al. (2018) , the recombination map was used in an QTL^{epi} mapping strategy. In this approach, the segregating parental DMRs were used as the predicting markers. The study identified genome-wide significant QTLs that predicted the performance of the hybrids in terms of flowering time, leaf area, and height. However, the authors noted that to gain a more detailed insight into the molecular mechanisms underlying the QTL effects, further studies would be required to create more epiHybrid families. This would allow to systematically investigate how parental methylation differences regulate hybrid methylome remodeling both locally (in *cis*) and in distal regions (in *trans*).

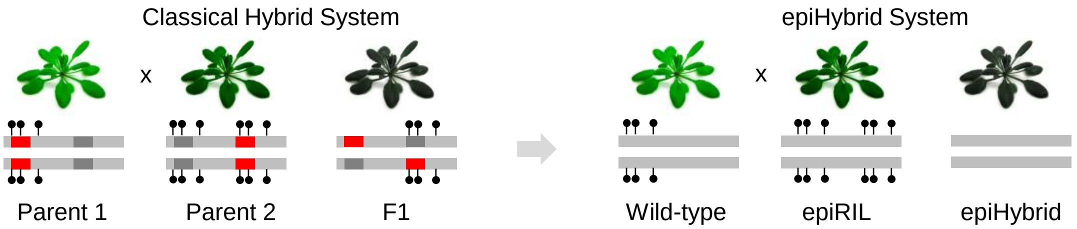


Figure 1.5: a) Crossing scheme explaining methylome remodeling in the F1 Hybrids (left) and in the epiHybrids (right). The two inbred parental lines, parent 1 and parent 2, differ both in fixed deleterious alleles and in their DNA methylation profiles (black lollipos). In F1 epiHybrids, the two parental lines are virtually isogenic but differ at the DNA methylation level. This experimental set-up can therefore be used to quantify epigenetic contributions to phenotypic heterosis and functional remodeling in their hybrid progeny. Red bars indicate deleterious alleles, while dark-grey bars indicate wild-type alleles.

These findings indicate that epigenetic markers can offer predictive insights beyond DNA sequence variation. Therefore, epigenetic measurements could complement genetic markers in crop systems where parental lines are highly polymorphic. For example, Seifert et al. (2018a,b) analyzed differences in sRNA expression patterns among heterotic groups in maize, which served as a strong predictor of grain yield in the hybrids. The predictive power of epigenetic markers was even greater than that of genetic markers in this study (Seifert et al. 2018a,b).

Beyond statistical predictions, gaining a better understanding of how epigenome diversity impacts plant phenotypes can provide valuable insights into the molecular regulation of agriculturally important traits. This knowledge has the potential to yield useful markers for breeding, reducing the need for extensive field trials and ultimately leading to increased crop production.

1.8 Aim of the project

Heterosis, commonly referred to as hybrid vigor, is a phenomenon in which the F1 hybrid offspring displays a superior phenotype compared to its parents. This concept has been extensively studied in agricultural breeding for decades and has contributed significantly to the improvement of crop performance. Most research on heterosis has taken a genetic approach to understanding how the interaction between different parental DNA sequence alleles can result in the heterotic phenotype observed in hybrid offspring. These studies propose that hybrid heterosis is the outcome of the complementation of deleterious recessive alleles that have become fixed in each parental variety and are brought together in the hybrid (Figure 1.5).

However, genetic explanations have often proven insufficient in explaining and predicting hybrid heterosis. Recent studies have demonstrated the critical role of epigenetic factors in this phenomenon. This is supported by evidence showing that F1 hybrids undergo significant DNA methylation remodeling relative to their parents, and experimental manipulations of DNA

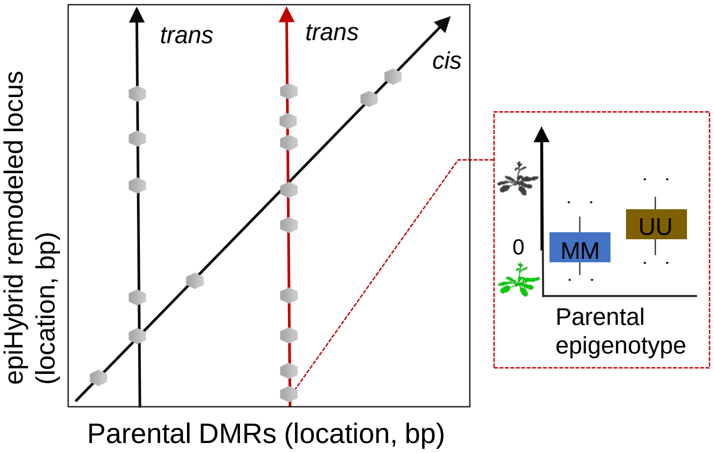

trans
trans
cis
epiHybrid remodeled locus (location, bp)
Parental DMRs (location, bp)
0
MM
UU
Parental epigenotype

Figure 1.6: The epiHybrid experimental system will give an insight into how parental differentially methylated regions (DMRs) (x-axis) predict methylome remodelling in the hybrids bothlocally and in distal regions. We hypothesize that highly pleiotropic parental DMRs contribute to phenotypic heterosis (right: boxplots) (Kakoulidou et al. 2024).

methylation pathways can affect the expression of heterotic phenotypes.

In most studies, the distinction between the genetic and epigenetic effects is unclear because the parental lines used for the hybrid crosses differ in both their genetic and epigenetic profiles (Figure 1.5). A significant challenge is to study the connection between parental and hybrid epigenomes independently of genetic variation and attribute any phenotypic changes in the hybrids to the epigenetic state of the parents, rather than DNA sequence polymorphisms.

To address this question, we will analyze a large experimental system of F1 epiHybrids. These epiHybrids were obtained by crossing a sterile Col-wt as the maternal parent with selected hypomethylated epiRILs as the paternal parents. The parents in this system are essentially isogenic but highly variable in their DNA methylation patterns. As a result, each epiHybrid inherits one wild-type chromosome from the recurrent female parent and one epiRIL hypomethylated chromosome from their respective paternal parents. Importantly, since the parental lines have nearly identical DNA sequences, the different epiHybrid populations differ only based on their paternal methylome contributions.

Hybrids of such crosses have displayed significant heterosis (Lauss et al. 2018), which indicates that experimentally-induced DNA methylation changes between the parents is sufficient to cause heterosis, independently of DNA sequence determinants. But to what extend particular features of parental epigenomes contribute to the widespread remodeling that occurs in the hybrids and to which extent the remodeling contribute to phenotypic heterosis is still unknown (Figure 1.5).

Importantly, we will attempt to associate local patterns of methylome remodeling found in the epiHybrids with parental DMRs using an QTL^{epi} mapping approach (Cortijo et al. 2014). Epigenetic mapping strategies can help to identify specific regions of the genome that are differentially methylated between parents and hybrids, and to link these regions with changes in gene expression and phenotype. This is similar to an expression QTL analysis and will yield associations both locally and in distal regions (Figure 1.6). We hypothesize that highly pleiotropic parental DMRs contribute to phenotypic heterosis observed in the epiHybrids and since we know for each marker the parental epigenotype of the paternal epiRIL, we can further assessed to which extent heterotic effects can be predicted based on specific DMRs.

The causal relationship between changes in the methylome and transcript levels and hybrid performance is a critical aspect of understanding heterosis. By integrating data from different omics, such as genomics, transcriptomics, and epigenomics, we will attempt to address the causal relationship between the observed changes in methylome and transcript levels and hybrid performance.

Chapter 2

Materials and Methods

2.1 Plant material

We obtained the *ddm1-2* epiRIL population (Johannes et al. 2009) from the Versailles Arabidopsis Stock center of INRA. They epiRIL lines were initiated by crossing two closely related parents of the same accession (Columbia, Col). The paternal parent was homozygous for the wild-type allele in the DDM1 and the maternal parent was homozygous for the *ddm1-2* mutant allele (Col-ddm1). Their F1 offspring were backcrossed as the maternal parent to the Col-wt and only the progeny plants containing the wildtype DDM1 allele were selected and propagated further for 6 rounds of single seed descent. The resulting epiRIL lines have very similar genomes but distinct epigenomes as the *ddm1-2*-induced DNA methylation differences were stably inherited (Johannes et al. 2009).

For the crosses, we selected a male-sterile Col-wt wild-type line as the maternal plant. Specifically, we used the msCol line, which was generated by crossing a Col-wt with the male-sterile line N75 (Melchinger et al. 2007) . The N75 line contains a recessive mutation in the male sterility 1 (MS1) gene, which controls the development of anthers and pollen in Arabidopsis. Homozygous mutants for the *ms1* allele are unable to produce viable pollen, resulting in male-sterile plants.

Although the MS1 gene was originally in the Ler background, Melchinger et al. (2007) introduced the mutant allele into a Columbia background by backcrossing it to a Col-wt for six generations and using marker-assisted selection to recover the recurrent Col-wt parent genome. The resulting msCol line contains both homozygous mutants and heterozygous seeds, and we obtained the msCol seeds from the University of Hohenheim.

To ensure genetic stability, we backcrossed the msCol progeny an additional three times at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK). Finally, to minimize errors during hand pollination and prevent self-fertilization, we used sterile seeds from two msCol plants of the same progeny, namely msCol-12 and msCol-16, as the maternal plants for the crosses.

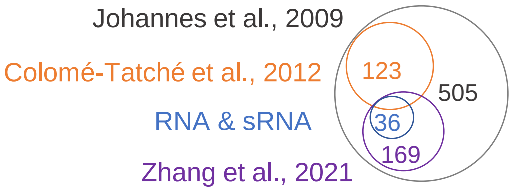


Figure 2.1: A Venn diagram summarizing how the paternal epiRILs used in this study relate to the epiRILs used in previous studies. The original epiRIL population (Johannes et al. 2009), the study that used tiling array data from 123 epiRILs and identified DMRs (Colomé-Tatché et al. 2012), the 169 epiRILs lines that were generated for the purposes of this study (Zhang et al. 2021) and the 36 epiRILs that were sequenced for sRNA and RNA expression.

2.2 Plant cultivation and crosses

To eliminate the chances that differences in the maternal cytoplasm would affect the phenotypes in the hybrids, we used the msCol plant as the female parent while the *ddm1-2* epiRILs were used as the male parents. We cultivated all parental plants in single seed pots in a greenhouse at the IPK. Pots were randomised and conditions were controlled for long day exposure (16h light, 8h dark), temperature, and humidity (20°C and 60 %).

Since the msCol recessive mutation is maintained in the heterozygous state in the progeny, only 50% of the seeds that we planted would have the recessive homozygous mutation. In order to use only those recessive sterile plants, we selected them at the inflorescence stage as they formed no pollen. As the first inflorescence matured, we examined the flowers under a microscope. If a msCol plant contained viable pollen, it was immediately removed from the greenhouse to avoid cross fertilisation. When the inflorescence of the maternal msCol plants had flowers opened and the stigmata were recessive, we used the epiRILs to fertilise the plants. One epiRIL paternal plant would fertilise one individual msCol plant. The pollinated female msCol inflorescence created the corresponding epiHybrid progeny, while the paternal epiRIL plants that were used for the cross were left with six siliques to dry and were used later as the male epiRIL parents in the phenotypic phase. In addition, we used msCol plants at the heterozygous state to pollinate one homozygous sterile plant and their resulting progeny was used later as the maternal parent in the phenotypic experiment.

2.3 High-throughput phenotyping

Phenotyping and imaging were conducted at the IPK phenotyping facility for small plants using mobile carriers covered with black rubber mats (Junker et al. 2015). Two phenotyping phases were carried out, each consisting of three individual replication experiments.

In the first phenotyping phase, we grew 382 epiRIL hybrids, their corresponding 382 paternal epiRILs, and the recurrent maternal msCol lines in parallel. For each cultivation experiment, six individuals per line were grown and phenotyped, resulting in 4,608 individuals per experiment

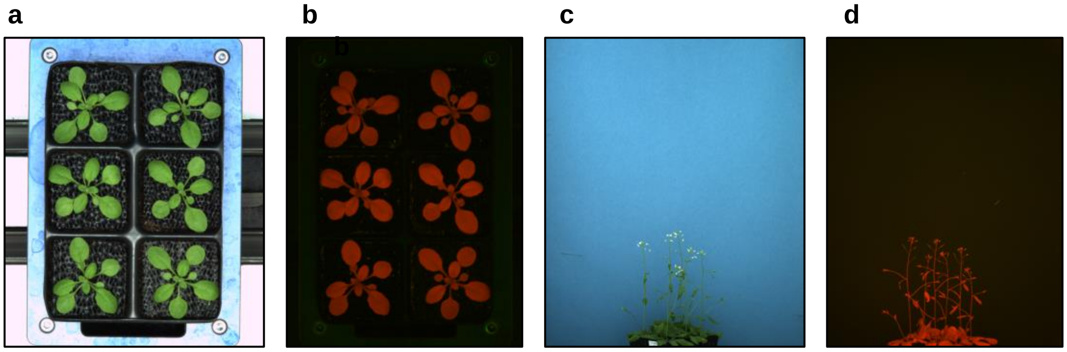


Figure 2.2: Images of the carriers inside the growth facility. Pictures ”a” and ”c” represents visible light imaging, while ”b” and ”d” represents fluorescence light.

and a total of 13,824 individuals across the three experiments. Seeds were sown randomly in 12-well trays, and the day of sowing was designated as day 0 (DAS). The pots were chilled for three days at 4°C in the dark and then acclimated for two days at 14−16°C under reduced light conditions before being grown under long-day conditions (16 hours of light, 8 hours of darkness) 20°C (18°C at night), with 60-75% humidity and 180−2401 μE light intensity for 20 days.

To ensure a diverse representation of phenotypic effects, we randomly selected 190 lines from the pool of epiRIL hybrids based on their leaf area measurements on DAS 18, excluding those previously used in publications (Figure 2.1). In the second phase, we performed three additional cultivation experiments, each consisting of the 190 selected epiHybrids grown in parallel with their corresponding 190 paternal epiRILs and the recurrent maternal msCol-wt line. As before, six individuals per line were grown and phenotyped in each experiment, resulting in a total of 6,912 individuals across the three experiments. The plants were grown in six-well trays until 27 days after sowing and evaluated for vegetative growth and developmental traits.

2.4 Imaging and image analysis

We used a LemnaTec-based system specifically designed for small plants, such as Arabidopsis, to capture images throughout the growth period. The system was set up in a growth chamber where the pots could move automatically on lanes. Once per day, the pots were placed in a side-small camera box, where the system captured automated images of the plants from different angles, providing top and side views in the visible (RGB) and near-infrared (NIR) wavelength range (Junker et al. 2015) (Figure 2.2). Additionally, the system allowed for imaging of static fluorescence and functional (kinetic) chlorophyll fluorescence analyses, as well as 3D laser scanning.

We utilized the Integrated Analysis Platform (IAP) open-source software to perform high-throughput plant image analyses (Klukas et al. 2014). From the images obtained, we extracted image-based plant traits using IAP, and measured leaf area on DAS 18 as the projected plant area in pixels under fluorescence light in order to study heterosis. To ensure consistency and accuracy

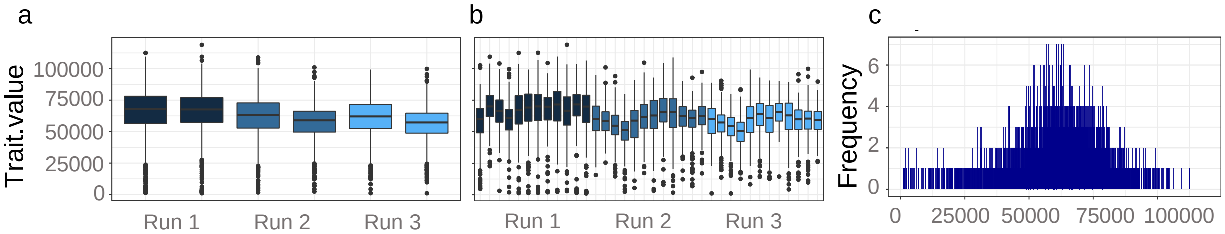


Figure 2.3: Leaf area across each replicate (a) and across each lane (b). c) Histogram of the normalized trait.

in data management, we linked the phenotyping workflows, plant growth imaging, and image analysis to a centralized data management system. All resulting data were stored and connected to a standardized metadata description in the management system, which helped streamline data processing and analysis.

2.5 Quantitative phenotypic analysis of heterosis

The high-throughput phenotyping data produced by the LemnaTec-based system were used to study heterosis. We normalized the raw leaf area measurements of each trait by removing outliers (>3SD) (Figure 2.3) and using the following mixed model in R (lme4) we accounted for environmental variation.

$$Y = Replicate + DtoG + Experiment + Block \tag{2.1}$$

Y denotes the phenotypic value of a trait and as fixed factor effects, we set germination date (DtoG), replicates within the experiments (Replicate), set of experiments (Experiment), and blocks of eight carriers moving together always in the phenotypic facility (Block). Each line (epiRIL or epiHybrid) that had fewer than 5 individual plants across the experimental replications was excluded from the analysis. Heterosis was computed using the output model residuals, in a likelihood approach (Lauss et al. 2018) and was defined as the epiHybrid's percentage change from the average of both parents (MPV).

$$\frac{epiHybrid - MPV}{MPV} \cdot 100 \tag{2.2}$$

2.6 DNA and RNA sample preparation

During the second phenotypic phase, all individual plants were harvested within a 3-hour time frame at 27 DAS. After removing the flowering stems and roots, the 18 individual rosettes from each line were pooled in 50 mL tubes. The samples were immediately frozen in liquid nitrogen

Table 2.1: Example of alignment and coverage summarising statistics

Line	Average sequencing	Genome coverage (%)
2F	22.3282	98.20187
2R	18.284	98.1987
4F	17.89	98.1042
4R	18.814	98.1081
9F	17.135	97.8966
9R	17.3788	98.0357
10F	22.3817	98.2673
10R	22.9986	98.2621

and stored at −80 °C until further processing. Genomic DNA was extracted from each pooled sample per accession using the DNAeasy plant mini kit from Qiagen, following the manufacturer's instructions.

To determine DNA concentration and purity, we used a 1 μL sample and a NanoDrop. At least 1 μg of DNA sample from each of the 169 epiHybrids, their corresponding 169 epiRIL paternal parents, and the 2 msCol maternal parents was sent to the Beijing Genome Institute (BGI), where it was prepared for library construction and whole-genome bisulfite sequencing (WGBS). Sequencing was carried out on an Illumina HiSeq X ten instrument, and clean raw paired-end reads (150 bp paired-end reads) were obtained for downstream analysis.

In addition, we selected 36 families, consisting of 36 epiRILs, 36 epiHybrids, and 2 msCol, for expression and small RNA (sRNA) analysis based on their performance during the second phenotypic phase. Specifically, we selected 12 families that showed high positive heterosis for leaf area, 12 with low parent heterosis, and 12 with no detected heterosis.

Total RNA was extracted using the miRNeasy kit from Qiagen, following the manufacturer's instructions. The extracted samples were sent to BGI for transcriptome and DNBSEQ UMI small RNA library preparation. Sequencing was performed on the DNBSEQ platform, and clean raw files were used for downstream analysis

2.7 Characterising hybrid remodeling

Methylstar (Shahryary et al. 2020a), as described by Shahryary et al. (2020a), was utilized to analyze the WGBS data for all lines. The resulting METHimpute files were then used downstream to call methylation regions per context, including CG, CHG, and CHH. Table 2.1 provides a summary of the alignment and coverage statistics, and the complete table can be found in Kakoulidou et al. (2024).

Using jDMR (Hazarika et al. 2022) the methylome was divided into sliding 200 bp regions (step size 50 bp) and only bins with at least 10 cytosines were retained. For each epiHybrid line, each 200 bp region was categorized as methylated, unmethylated, or intermediate methylated, while the parental lines were classified as methylated or unmethylated. These results were saved in a binary matrix file with the state calls across all lines. “0” indicates an unmethylated region, “1” a methylated region and “0.5” an intermediate methylation state (Table 2.2).

To determine whether the DNA methylation state of each epiHybrid differed from the expected midparent value (MPV; ((msCol + epiRIL)/2)) of its two parents, we quantified, for each 200 bp region, the methylation state based on the state calls. Parental divergence (PD) was computed by subtracting the methylation state call of the epiRIL from that of the msCol, while HD was calculated by subtracting the MPV from the methylation state call of the F1 hybrid. Non-additivity (NAD) was defined as HD differing from zero, while additivity equaled zero.

Regions were labeled differentially methylated regions (DMRs) or similarly methylated regions (SMRs) based on PD values. A region was called a DMR when the PD differed from zero and an SMR when PD equaled zero (Table 2.2).

In addition to the binary matrix file, an output ”rc.meth.lvl.txt” matrix was created than contains the methylation levels for each sample at each specific region. Using the methylation levels we calculated the epiHybrid’s divergence change using the formula:

$$\frac{epiHybrid - MPV}{MPV} \cdot 100 \tag{2.3}$$

2.8 RNA-seq analysis

We performed quality checks on the raw reads using FastQC, removed low-quality reads using Trim Galore, and aligned the remaining reads to the reference *A. Thaliana* (TAIR10) genome using Tophat2 (Kim et al. 2013). We then used featureCounts (Liao et al. 2014) to count the reads, and the resulting raw count table was processed in R.

To filter out genes with low expression levels, we retained reads for a given gene only if it had at least 1cpm (counts per million) in at least two samples. We then used the trimmed mean of M-values (TMM) method in the edgeR package (Robinson et al. 2010) to normalize the reads. We calculated the expression divergence of the epiHybrids using the resulting normalized counts, according to Eq. 2.3.

2.9 sRNA-seq analysis

We processed the sRNA data using Shortstack (Johnson et al. 2016) with default parameters. Specifically, we provided the path to the raw data file using the ”readfile” parameter, specified the output directory name using ”outdir,” provided the reference genome using ”genomefile,” disabled the MIRNA search using ”nohp,” and specified the genomic intervals to be analyzed using ”locifile.” We used 200 bp regions for each context separately as the intervals. For each line, the resulting output files contained raw read counts for each sRNA size length. We normalized these counts using the TMM method from the edgeR package (Robinson et al. 2010) and used them to calculate the epiHybrid’s sRNA divergence using the formula in Eq. 2.3 downstream.

Table 2.2: Remodeling senario at each 200bp region

msCol	epiRIL	PD	DMR/SMR	MPV	epiHybrid	HD	Scenario	NAD/ADD
1 (M)	1 (M)	0	SMR	1	1 (M)	0	A	ADD
1 (M)	1 (M)	0	SMR	1	0.5 (I)	−0.5	B	NAD
1 (M)	1 (M)	0	SMR	1	0 (U)	−1	C	NAD
0 (U)	0 (U)	0	SMR	0	1 (M)	1	D	NAD
0 (U)	0 (U)	0	SMR	0	0.5 (I)	0.5	E	NAD
0 (U)	0 (U)	0	SMR	0	0 (U)	0	A	ADD
1 (M)	0 (U)	1	DMR	0.5	1 (M)	0.5	F	NAD
1 (M)	0 (U)	1	DMR	0.5	0.5 (I)	0	G	ADD
1 (M)	0 (U)	1	DMR	0.5	O (U)	−0.5	H	NAD
0 (U)	1 (M)	−1	DMR	0.5	1 (M)	0.5	I	NAD
0 (U)	1 (M)	−1	DMR	0.5	0.5 (I)	0	J	ADD
0 (U)	1 (M)	−1	DMR	0.5	0 (U)	−1	K	NAD

2.10 DNA-seq analysis

To extract genomic DNA from the two msCol plants, we followed the same protocol as for the WGBS samples, and sequenced them at BGI using an Illumina HiSeq X ten. We performed quality control of the raw reads with FastQC (Andrews 2010) and quality-trimmed them in paired-end mode using Trimmomatic (Bolger et al. 2014) with the following parameters: LEADING:5, TRAILING:5, MINLEN:50, SLIDINGWINDOW:3:18. We then mapped the trimmed reads to the reference *A. Thaliana* (TAIR10) genome using Bowtie2 (Langmead and Salzberg 2012). We marked duplicate reads using Picard and sorted them with Samtools (Danecek et al. 2021).

To identify polymorphisms between the Columbia ecotype genome, we used the ”Haplotype-Caller” tool of GATK with default parameters(Auwera and O’Connor 2020) for variant calling. We extracted single nucleotide polymorphisms (SNPs) and INDELs separately using GATK’s ”SelectVariants” tool, removed low-quality reads and heterozygotes (QUAL > 40), and performed base quality score recalibration. We then performed a second round of variant calling using the recalibrated bam file and annotated the filtered variants with snpEff (Cingolani et al. 2012). All analysis steps were performed separately for the two msCol siblings (msCol12 and msCol16). We used the ”intersect” tool from Bedtools (Quinlan and Hall 2010) to find the common overlap between the two lines.

2.11 DMR detection between the msCol plants and the Col-wt

To confirm that the msCol methylome closely resembles that of the Col-wt, we compared the two genomes for DMRs. First, we identified DMRs between the msCol and a publicly available Col-wt plant, the Col0 GO MA3 line (GEO accession number GSE153055) (Shahryary et al. 2020b). To ensure that the number of DMRs detected between the msCol and Col-wt was similar to the number between two Col-wt plants, we also analyzed another published Col-wt line (Col-wt replicate 1, GEO accession number GSE70912) (Yang et al. 2016) for DMRs between the Col0 GO MA3 line and the Col-wt replicate 1. We used jDMR (Hazarika et al. 2022) to identify these regions.

2.12 Construction of an epiRIL linkage map

In a previous study, we used tiling-array data from 123 epiRILs and their *ddm1-2* and Col-wt founder lines to identify DMRs in the parental lines that were stably inherited in the epiRILs through at least 10 rounds of meiosis. Using these stable DMRs as physical markers, we constructed a linkage map (Colomé-Tatché et al. 2012).

In this study, we utilized the methylomes of 169 epiRILs to verify and improve map resolution. We began with the same tiling-array derived linkage map of the 126 DMRs, but found that only 37 families overlapped with the epiRILs in our study. For these overlapping families, we verified that the methylation state calls produced from our WGBS at the same DMR positions were consistent with the tiling-array calls. To improve map resolution, we created a new linkage map with 144 markers that had better coverage in the chromosome arms. For each DMR marker, a given epiRIL paternal parent is either epihomozygous for the wild-type methylated state (MM) or epihomozygous for the *ddm1*-like state (UU), as described in Zhang et al. (2021).

We made use of these 144 stable segregating DMRs as markers in a QTL^{epi} mapping analysis as was described in Cortijo et al. (2014). The DMR markers were used as predictors and the change of hybrid divergence (Eq. 2.3) at each remodeling region in the epiHybrids was used as the molecular quantitative trait. Only remodeling regions that showed a remodeling TCM or TCdM event in more than 50% of the epiHybrid families were selected. Hybrid divergence (Eq. 2.3) was normalized with the Ordered Quantile (ORQ) normalization transformation and the values were used in an QTL^{epi} analysis as described previously (Cortijo et al. 2014; Broman et al. 2003; Kakoulidou et al. 2024).

The extent to which NAD-QTLs can account for the variance in mid-parental methylation divergence is quantified by the coefficient of determination (R^2), while the direction of the effect is given by the regression slope (a $>$ 0 for a positive effect or a $<$ 0 for a negative effect). A positive effect indicates that when the parental lines are differentially methylated, their hybrid offspring exhibit a greater mid-parental methylation divergence compared to the parents that are homozygous methylated. Conversely, a negative effect suggests that differential methylation between the parental lines results in a reduced mid-parental methylation divergence compared to the parents that are homozygous methylated."

To distinguish local from dista-acting effects we obtained for each NAD-QTL^{epi} the confidence intervals (CI) around the peak QTL position using a 2 LOD drop-off criterion. If the body of the NAD target was located within a given CI, the NAD-QTL was defined as locally-acting (in *cis*), otherwise as distally-acting (in *trans*).

2.13 Annotation enrichment analysis

The jDMR package's "annotateDMR()" function (Hazarika et al. 2022) was used to annotate regions of interest, using updated annotation files for genes and TEs downloaded from Ensembl Plants (http://plants.ensembl.org/info/data/ftp). The list of gbM genes was obtained from Zhang et al. (2021), while the list of 24-nucleotide sRNAs was obtained from Slotkin et al. (unpublished data). The resulting gene list was analyzed using ShinyGo (Ge et al. 2020) to gain insights into their gene ontology.

2.14 Epigenome-wide Association Study (EWAS)

To explain the total phenotypic variance attributed to the detected QTL^{epi}, we selected the peak QTL^{epi} marker for each chromosome that showed a statistically significant association with LA heterosis (ANOVA; $p\text{-value} > 0.05$). We then used the parental QTLepi genotypes as predictors and the HD of leaf area as the response variable to create a regression model, which we refer to as the core model. To investigate whether *de novo* NADs can explain heterotic variance beyond what is already accounted for by the core model, we conducted a conditional epigenome-wide association study (EWAS) by adding HD in any given de novo NAD as an additional predictor in the core model. *De novo* NADs are remodeling regions that did not show any association with parental DMRs in the NAD-QTL^{epi} analysis. To group the *de novo* NAD regions that exhibited a significant improvement in the core model, we performed Hierarchical Clustering on Principal Component Analysis (HCPCA) with a false discovery rate (FDR) less than 0.05, as determined by the Benjamini-Yekutieli method. For each cluster, we selected the de novo NAD region with the highest improvement in a single model as the *de novo* marker (Kakoulidou et al. 2024).

2.15 Statistical, Graphics, and Visual Analyses

All statistical analyses were perfomed in R (Team 2020). To visualize the genome-wide DNA methylation profiles of the epiRILs, we used the Integrative Genomics Viewer (IGV) software (Robinson et al. 2011). This enabled us to interactively explore the DNA methylation data and visually inspect methylation patterns across different genomic regions.

2.16 Data availability

Sequencing data produced for this study have been deposited in the NCBI GEO database under GSE211719.

Chapter 3

Results

3.1 Construction of a large epiHybrid population

To study how methylation remodeling in hybrids is driven in the absence of genetic differences we generated a panel of 500 *A. thaliana* F1 epiHybrid families by crossing a male sterile plant of the Columbia accession (msCol), to 500 different paternal *ddm1-2*-derived epiRILs (Johannes et al. 2009) (Figure 3.1). As the maternal plant for the crosses, we used a male sterile plant that has a mutation in the MALE STERILITY 1 (MS1) gene and does not produce viable pollen. This helped minimize the risk of errors during hand pollination or unwanted self-fertilization.

The MS1 allele was originally present in a Landsberg (Ler) genetic background and was subsequently transferred into the Col-wt genetic background through six generations of repeated backcrossing (Melchinger et al. 2007). To study the successful introgression of the MS1 allele, whole-genome re-sequencing of the msCol plants was performed. The results of this analysis revealed a homozygous 2.7 Mb introgressed Ler segment on the arm of chromosome 5 (7.36 – 7.37 Mb), encompassing the MS1 locus. Outside of this region, only a small number of homozygous, non-coding SNPs and small INDELs were detected (Figure 3.2).

To confirm the similarity of the msCol methylome and the Col-wt methylome, we identified DMRs between the two lines. Our analysis revealed that the msCol line has a highly comparable methylome to the publicly available Col-wt line (Figure 3.3). Thus, the introgression of the *ms1* allele did not result in any observable *trans* effects on DNA methylation patterns. Moreover, the two plants were also found to be highly similar at the phenotypic level (Figure 3.3c).

Although the DNA sequence background of the *ddm1-2*-derived epiRIL lines used as the paternal parent is almost identical to that of the Col-wt (Figure 3.4), their DNA methylation patterns were significantly different. This is due to the induced hypomethylation caused by the *ddm1* mutation, which resulted in the inheritance of many hypomethylated regions across their genomes (Johannes et al. 2009; Colomé-Tatché et al. 2012; Zhang et al. 2021).

In addition to the previously identified increased mobilization rates of certain TE families, further analysis of the epiRILs have revealed that the majority of the TE insertions were private to each epiRIL and did not have any observable phenotypic effects (Quadrana et al. 2019; Cortijo et al. 2014; Johannes et al. 2009; Marí-Ordóñez et al. 2013). This suggests that the TE insertions may not be contributing to the observed heterosis in the epiHybrids. Furthermore, analysis of single nucleotide polymorphisms (SNPs) and insertions/deletions (Indels) in the epiRILs did not reveal any significant differences compared to the Col-wt plant (Lauss et al. 2018). Therefore, the observed heterosis in the epiHybrids is likely due to the differential DNA methylation patterns inherited from the paternal parent, rather than genetic variation at the nucleotide level.

In our crossing experiment, both parental lines of each cross share the same DNA sequence as the Col-wt, but differ significantly in their DNA methylation levels. To eliminate any potential contribution of molecular or phenotypic variation from the maternal line, we used the msCol line as the recurrent mother across all crosses. This approach ensured that any phenotypic differences observed among the resulting F1 epiHybrids could only result from the paternal methylome contributions. By systematically examining how specific regions of the paternal methylomes contribute to heterosis, we aimed to gain a deeper understanding of the molecular mechanisms underlying this phenomenon.

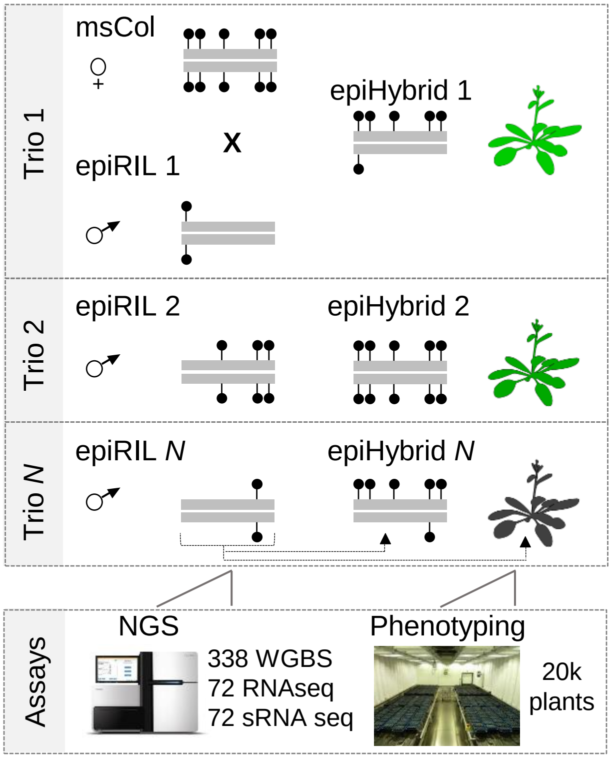


Figure 3.1: Overview of the construction of the epiHybrid population using the epiRILs as the paternal lines.We constructed the epiHybrid population by crossing each of the 382 epiRIL lines with a recurrent maternal msCol line. All plants were subjected to phenotyping, and plant material was collected for WGBS, RNA sequencing, and sRNA sequencing. Each trio consisted of a maternal msCol line, a specific epiRIL line, and the corresponding epiHybrid (Kakoulidou et al. 2024).

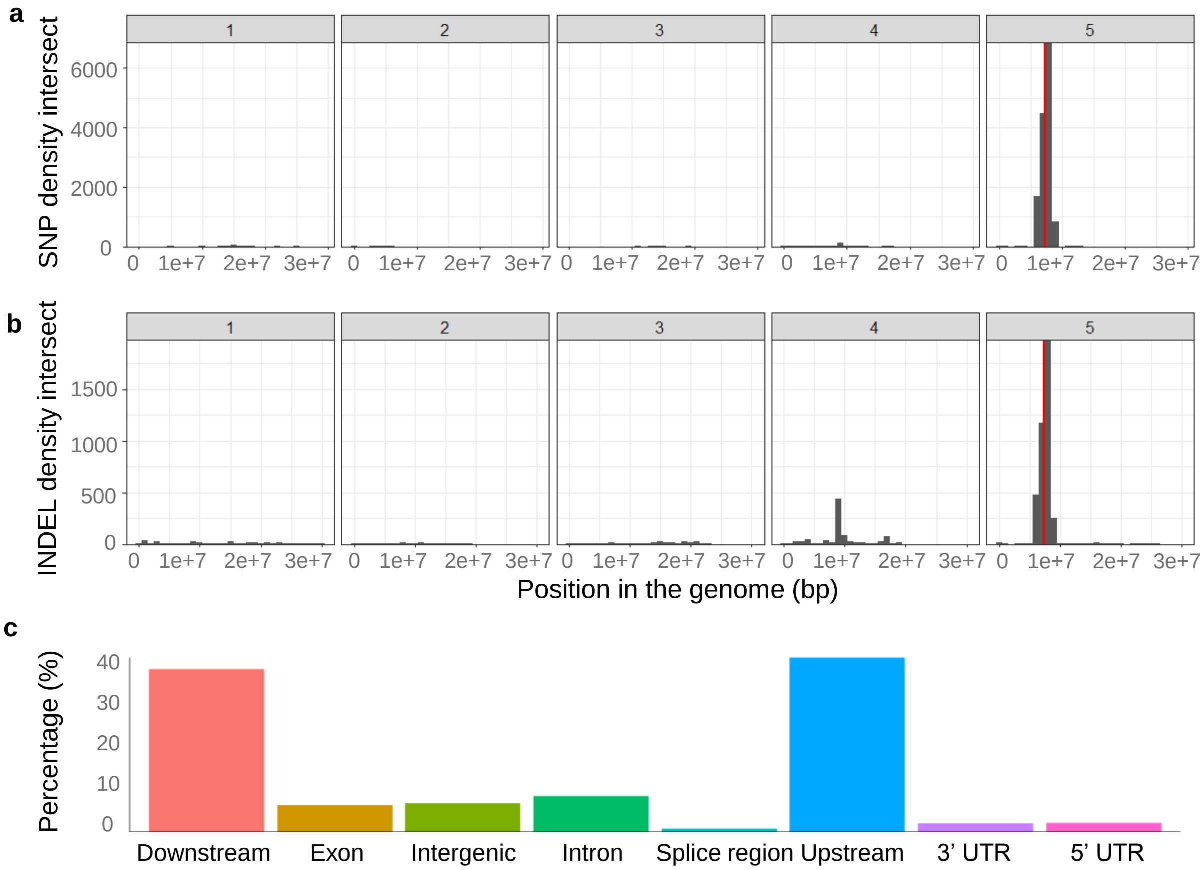


Figure 3.2: The genomes of two siblings, msCol12 and msCol16, are compared. a) The histograms depict the density of SNPs within the genomic intersection of msCol12 and msCol16. This provides a visualization of where SNPs are distributed across the genome in this region of interest. b) Another metric analyzed in this study is the density of INDELs within the same genomic intersection. The histogram shows how the frequency of INDELs compares to SNPs in this region of interest, with red lines indicating the location of the MS1 locus. c) The number of variants in the two siblings (msCol12 and msCol16) is represented in bar plots, with mean percentages displayed. These values were extracted from the summary output of snpEff (Cingolani et al. 2012), which provides information on the impact of genetic variants on protein function. These results can provide insights into the genetic differences between the two siblings and how these variations may affect their phenotype (Kakoulidou et al. 2024).

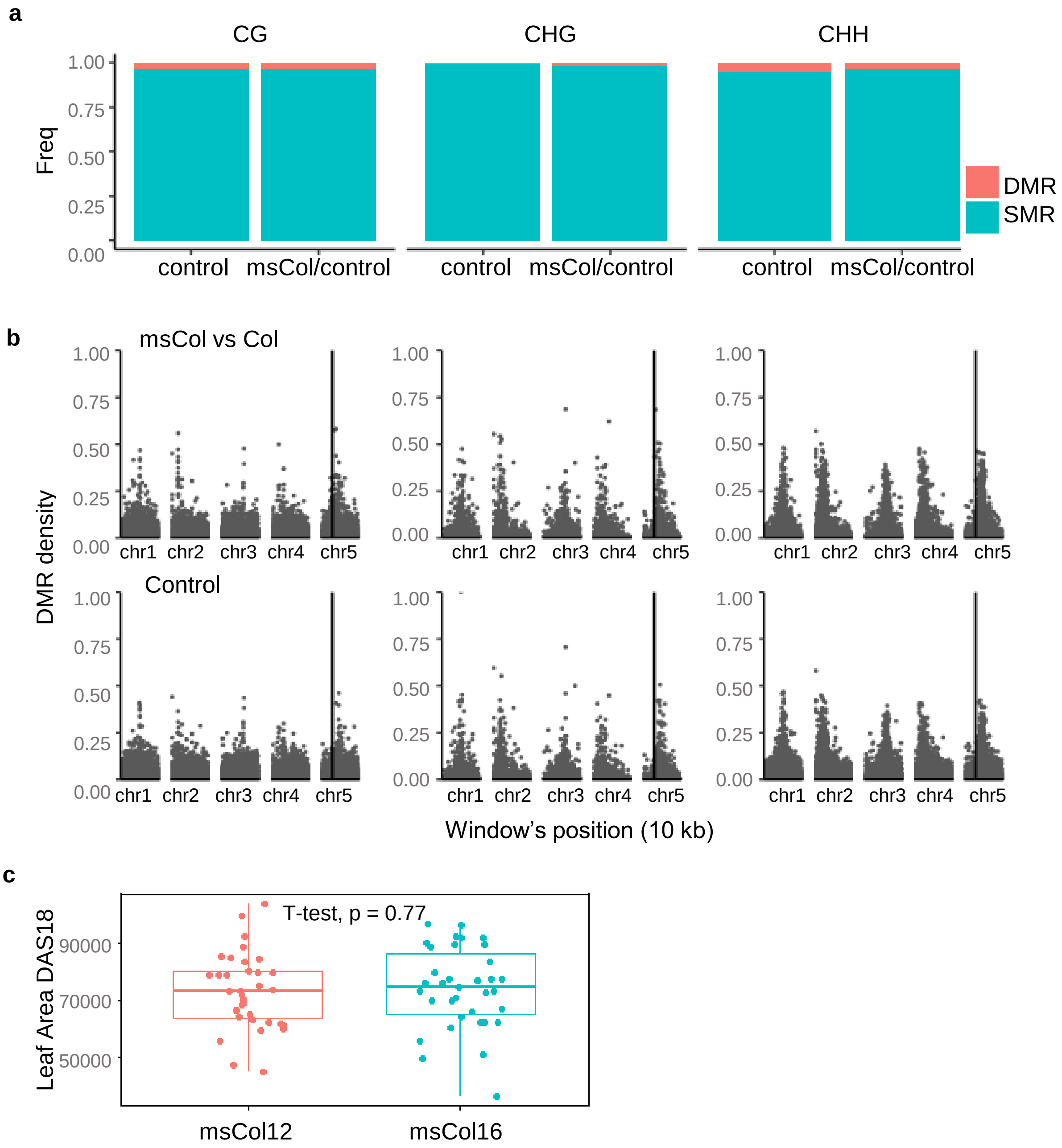


Figure 3.3: a) The stacked bar plot illustrates the frequency of DMRs in two datasets. Blue indicates the frequency of DMRs when comparing msCol with Col-wt, while red represents the frequency of DMRs when comparing two Col-wt lines. b) Histogram of DMR density across genome for both datasets, with MS1 locus (chromosome 5) indicated by black line (Kakoulidou et al. 2024). c) Leaf area compared for two siblings at 18 DAS in phenotypic phase 2.

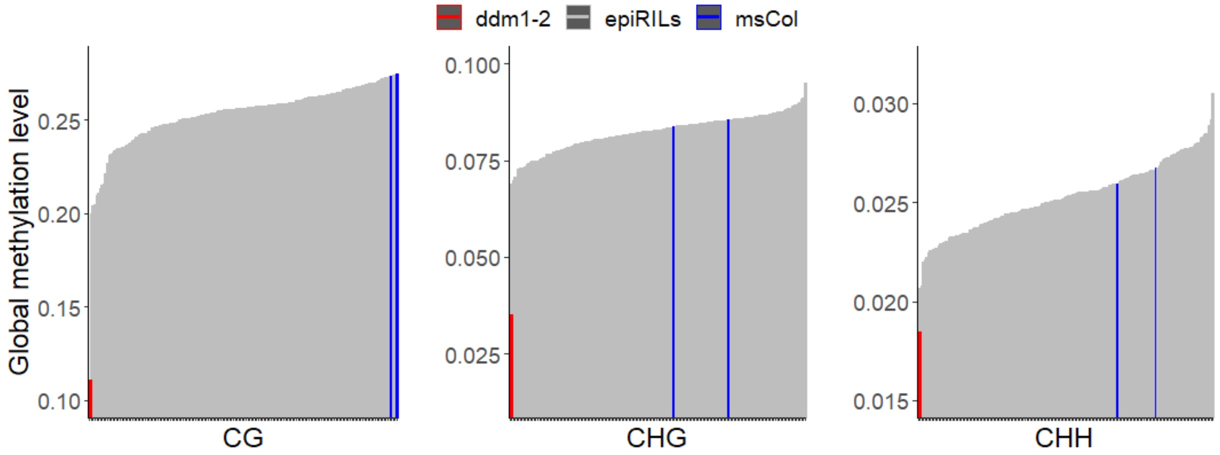


Figure 3.4: Grey bars represent the global methylation level for each paternal epiRIL, red bars represent the *ddm1-2* mutant, while blue bars indicate the msCol maternal lines (Kakoulidou et al. 2024).

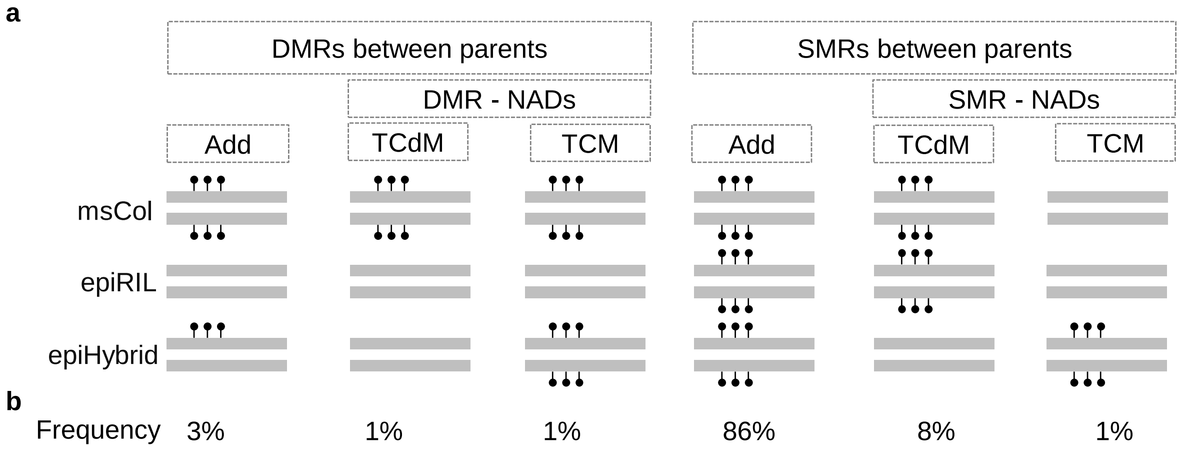

a
DMRs between parents
SMRs between parents
DMR - NADs
SMR - NADs
Add
TCdM
TCM
Add
TCdM
TCM
msCol
epiRIL
epiHybrid
b
Frequency
3%
1%
1%
86%
8%
1%

Figure 3.5: Schematic model that was used to characterize methylome remodeling regions.
a) Genome was divided in 200 bp regions and each of the regions was characterized either as differentially methylated (DMR) or similarly methylated (SMR) depending on the methylation status of the two parents. By comparing the region-level methylation states of the epiHybrids with those of the two parents, each region was classified as additive (ADD), trans-chromosomal methylated (TCM), or trans-Chromosomal de-methylated (TCdM). TCM and TCdM regions are collective refer to as non-additive regions (NAD). b) Genome-wide frequency of each remodeling category (Kakoulidou et al. 2024).

3.2 Patterns of local methylome remodelling in epiHybrids

To study how the methylome is being remodeled in the epiHybrids, we performed WGBS for 169 epiHybrids (pooled siblings) and their parental lines. We divided the genome into 200 bp regions (window size of 200 bp, step size of 50 bp) and we compared the methylation status (methylated or not methylated) of each epiHybrid to the methylation status of its two parents (Figure 3.5a).

In most of the regions (89%), the methylation status of the parents was stably inherited to the epiHybrids (Figure 3.7, Figure 3.5b). But consistent with previous findings (Rigal et al. 2016; Shen et al. 2012; Lauss et al. 2018; Dapp et al. 2015; Zhang et al. 2016), a considerable proportion of the regions (11%), showed non-additive methylation (NAD) in the epiHybrids, indicating that the methylation level of the F1 hybrids diverged from what would be expected if the methylation level of the parental alleles was stably inherited. By merging neighbouring regions together, we show that NAD regions have an average size of 200-300 bp. Browser shots of representative regions give a better overview of their methylation status (Figure 3.6). Methylome remodeling was most dominated (87%) in regions where DNA methylation status were similar between the epiRILs and the msCol (SMRs) and these regions were highly enriched for CHH sites and primarily within TEs (Figure 3.8a,b).

In contrast, a relatively small fraction (13%) of the remodeled regions were located in parental DMRs. Nonetheless, given that DMRs account for only 4% of the parental genome, this 13% represents a significant enrichment (bootstrap test, p-value < 0.0001)(Figure 3.9a). Interestingly, a larger proportion of parental DMRs (34%) were remodeled in the epiHybrids compared to only 10% of SMRs, indicating that DMRs are more prone to remodeling events.

As was mentioned above, a well-known DMR-associated remodelling mechanism is TCM (Greaves et al. 2012a). It involves an unexpected increase of local methylation levels in the hybrids and 24nt small RNAs that are initially produced from the methylated parental allele but are targeting the unmethylated allele for *de novo* methylation (Shivaprasad et al. 2012; Greaves et al. 2012a; Chandler 2007). To test the sRNA involvement in our TCM classified regions, we performed small RNA sequencing for 36 epiHybrids, their corresponding 36 epiRIL paternal lines and the 2 msCol maternal lines. Similarly to previous studies 21nt and 24nt sRNAs were the most abundant in the epiHybrids (Figure 3.9b).

The observed patterns in TCM-DMRs regions are in line with sRNA-based mechanisms as reported by Zhang et al. (2016). Specifically, we observe a non-additive gain in the local abundance of 24-sRNA in TCM-DMRs regions relative to the MPV. In some cases, the levels of 24-sRNA in these regions are equal to or even greater than those of the homozygous methylated parent, indicating a potential role of sRNAs in regulating DNA methylation (Figure 3.8). Conversely, in parental DMRs (TCdM-DMRs), we observe a reduction in sRNA production that is correlated with methylation loss.

We also checked if a similar correlation stands for the TCM remodeling events happening in SMRs. We hypothesized that such TCM-SMRs events are the result of TCM remodeling regions occurring in nearby DMRs. To explore this, we calculated the distance of each remodeling SMR (either TCM or TCdM) to the closest corresponding (TCM or TCdM) DMR-NAD (Figure 3.8d).

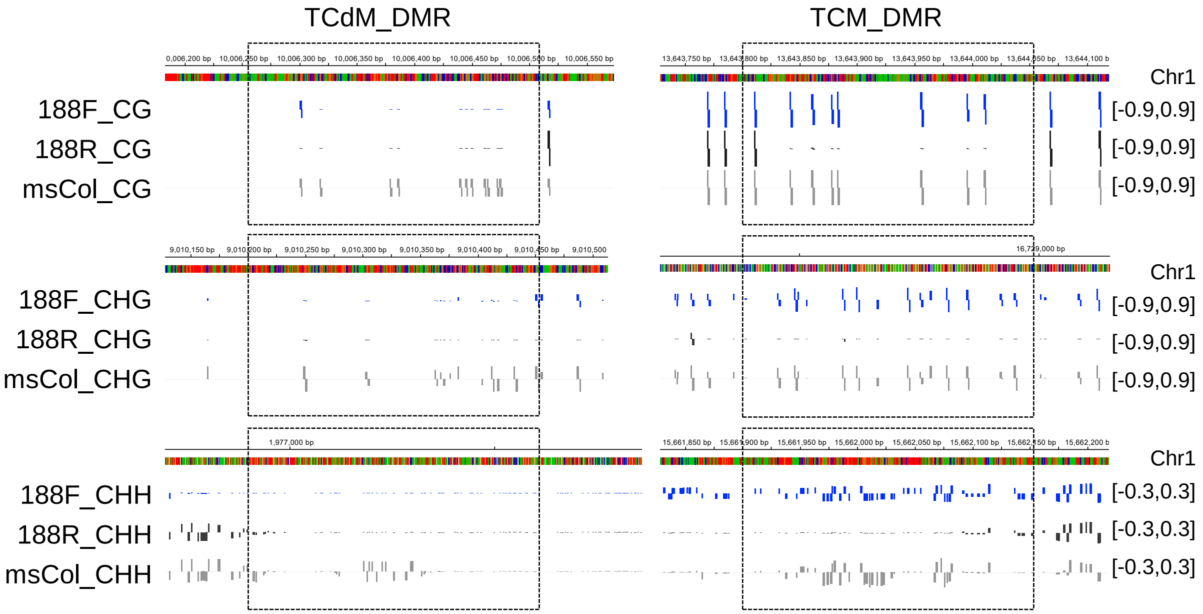


Figure 3.6: Representative browser shots of NAD regions for the 188 family for each context (Kakoulidou et al. 2024).

We found that on average TCM-SMRs are only about 350 bp away from TCM-DMRs, and display similar sRNA changes patterns. Since a typical 24nt sRNA cluster is only 918 bps in length (Figure 3.10), it could means that many TCM-SMR events are just the byproduct of sRNA changes at neighbouring TCM-DMRs.

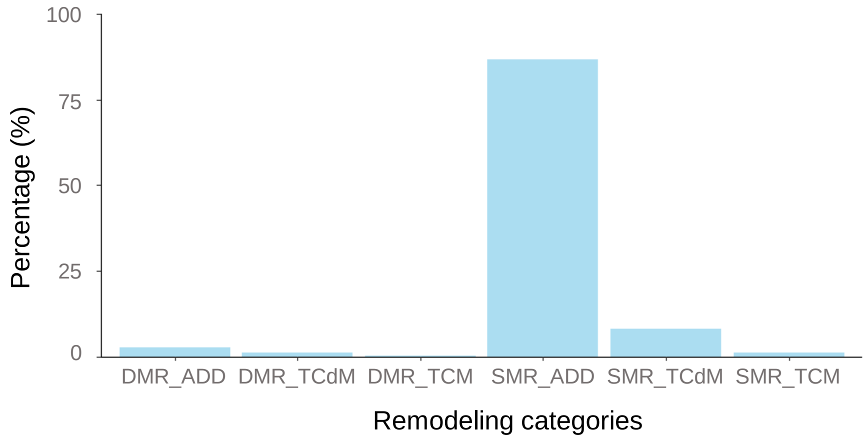
Similar conclusions could be obtained for NADs where the epiHybrids loose methylation while both parents are methylated (TCdM-SMRs). However, the trend was less profound, as the closest TCdM-DMR was relatively far away (1650 bp, on average) (Figure 3.8d), indicating that there is an additional mechanism affecting SMRs, possibly acting from a distal location (in *trans*).

3.3 Distal regions undergo widespread co-remodeling in epiHybrids

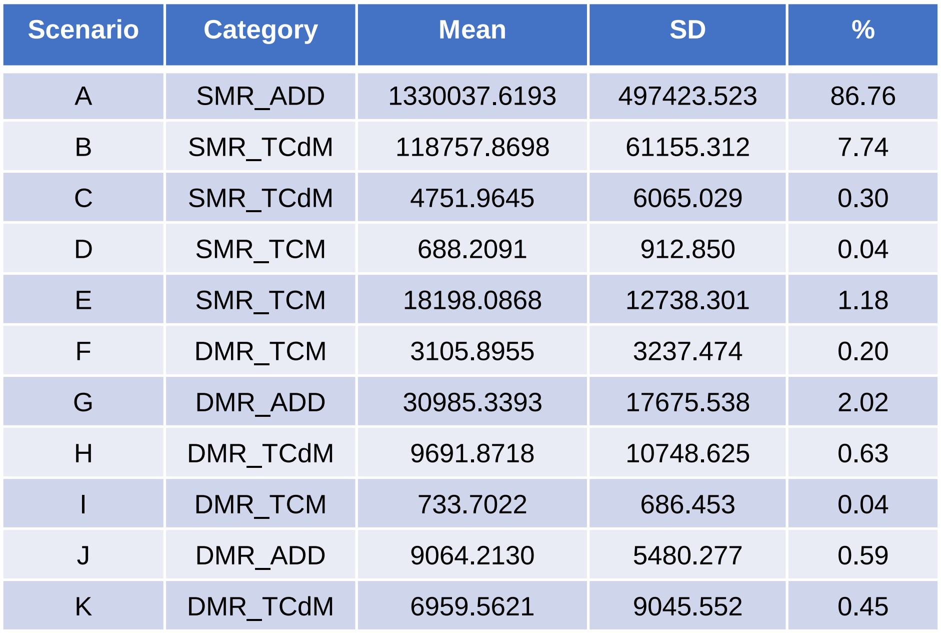
The enrichment in TEs is reflected in the genome-wide distribution of the remodeling regions, which show a clear density around pericentromeric regions of each chromosome (Figure 3.12a). Although this distribution pattern is consistent across epiHybrids, we observed significant variation in the frequency of NADs among epiHybrids, especially within pericentromeric regions.

Our experimental design, which involved multiple families-crosses, allowed us to investigate the correlation between NADs in any two regions of the genome and identify remodeling events that occurred simultaneously. This is a significant advantage compared to previous studies that focused on only a few families-crosses.

a



b



Scenario	Category	Mean	SD	%
A	SMR_ADD	1330037.6193	497423.523	86.76
B	SMR_TCdM	118757.8698	61155.312	7.74
C	SMR_TCdM	4751.9645	6065.029	0.30
D	SMR_TCM	688.2091	912.850	0.04
E	SMR_TCM	18198.0868	12738.301	1.18
F	DMR_TCM	3105.8955	3237.474	0.20
G	DMR_ADD	30985.3393	17675.538	2.02
H	DMR_TCdM	9691.8718	10748.625	0.63
I	DMR_TCM	733.7022	686.453	0.04
J	DMR_ADD	9064.2130	5480.277	0.59
K	DMR_TCdM	6959.5621	9045.552	0.45

Figure 3.7: Statistics for the methylome remodeling categories. a) The bar plot displays the distribution of remodeling categories in different regions, where NAD regions are included as TCM and TCdM events. The proportions of each category are depicted in the graph, providing a visual representation of the frequency of each remodeling event. b) The table presents detailed statistics for each remodeling category, including the number of regions and the percentage of the genome covered by each category (Kakoulidou et al. 2024).

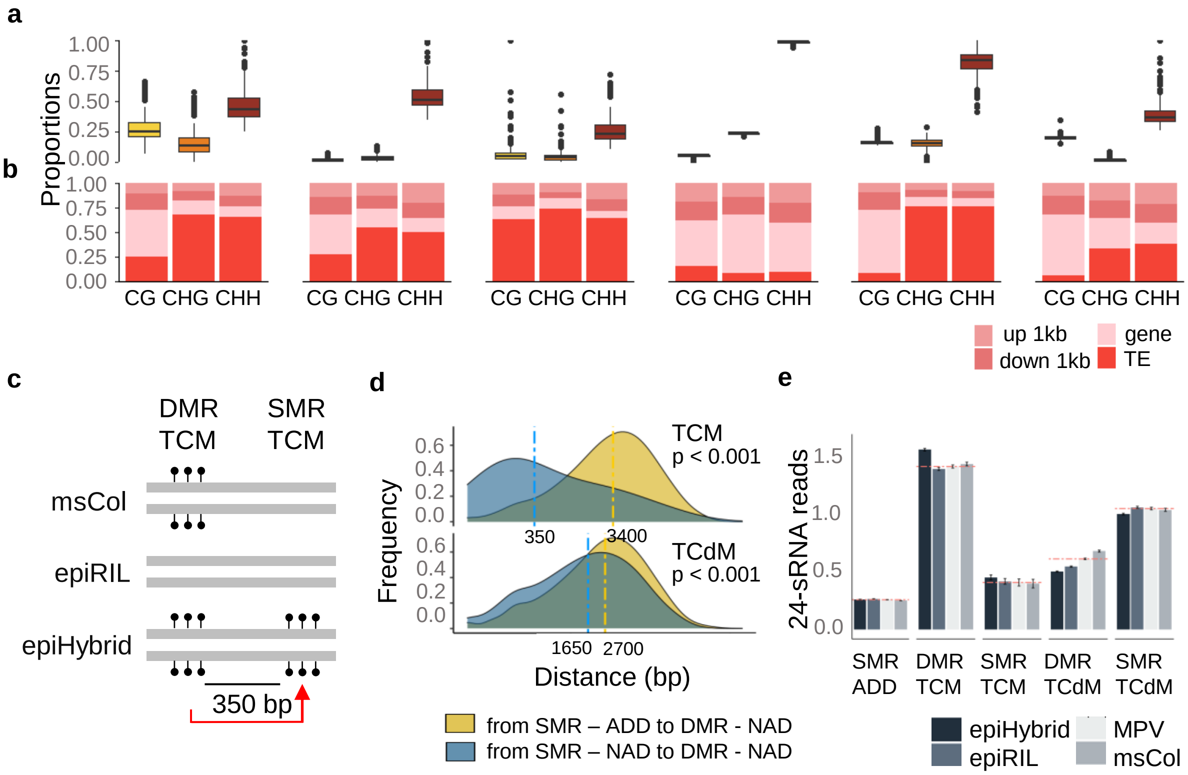


Figure 3.8: Mechanistic insights of the remodeling regions. a) The remodeling categories are displayed in a partitioned format according to cytosine context. b) The annotation enrichment for each remodeling category is illustrated in this figure. c) A schematic to visualize how a DMR-TCM event leads to a TCM event in a proximal SMR. d) This plot displays the frequency of distances for DMR-TCM and DMR-TCdM regions. In the upper plot, yellow indicates the distance of a given SMR-ADD region to a DMR-TCM region, while blue indicates the distance of a given SMR-TCM region to a DMR-TCM region. The lower plot shows yellow for the distance of a given SMR-ADD region to a DMR-TCdM region, and blue for the distance of a given SMR-TCdM region to a DMR-TCdM region. Vertical lines indicate the median for each group, with all P-values being significant. e) These barplots illustrate the normalized level of 24-sRNA reads for each category in the CHH context. The mean value of the middle-parental value (MPV) is indicated by a horizontal line (Kakoulidou et al. 2024).

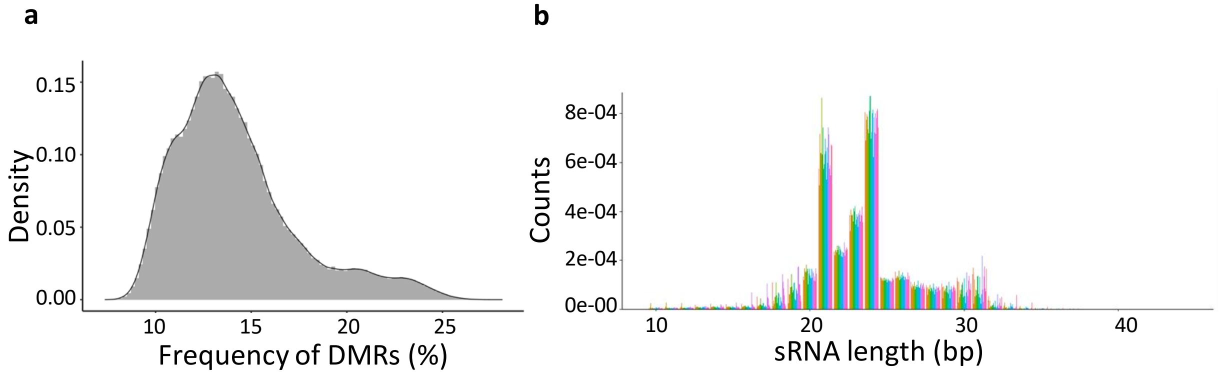


Figure 3.9: a) Density plot shows the DMR frequency distribution for all remodeling regions in bootstrap samples across 169 epiHybrid lines. We performed 1000 times a permutation test to randomly select remodeling regions and calculate the frequency of NAD-DMRs within selected NADs (Kakoulidou et al. 2024). b) Length distribution of sRNA sequencing reads in the 36 epiHybrids. Each color represents on family.

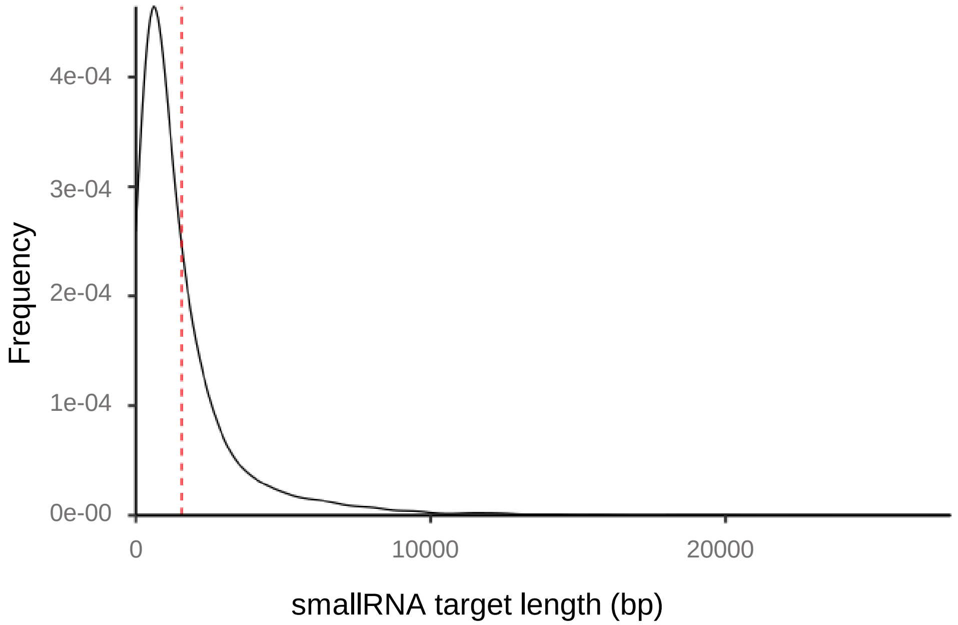


Figure 3.10: Length distribution of known sRNAs target regions (medial 918 bps) (Kakoulidou et al. 2024).

In order to gain insights into the frequency and patterns of remodeling events in different regions across a variety of epiHybrids, we performed an analysis to determine the occurrence of remodeling events in each region. Specifically, we examined the number of epiHybrids that exhibited remodeling events in each region, and classified regions accordingly. Regions where remodeling events were observed in only one epiHybrid were classified as ”private,” indicating that these events were unique to that particular epiHybrid. Conversely, regions with remodeling events observed in multiple epiHybrids were labeled as ”shared,” indicating that these events occurred across multiple epiHybrids (Figure 3.11).

A substantial proportion of NAD regions were found to be shared among different epiHybrid families, indicating that the occurrence of remodeling events is not random but rather a systematic feature of these genomes. However, a significant number of NAD regions were also identified as private regions, although they were much less frequent than the shared ones. Interestingly, shared NAD regions were found to be highly concentrated in pericentromeric regions, while private regions were more commonly observed in euchromatin (Figure 3.12b). To focus on the most robust remodeling events, we specifically analyzed NADs that were shared by at least 10 epiHybrid families. This revealed that a substantial proportion of NADs (21% for CG, 15% for CHG, and 23% for CHH) were consistently remodeled across different epiHybrids, indicating that local remodelling events are a reproducible, rather than a random, feature of the epiHybrid genomes.

Using these shared NADs, we treated the percent of mid-parent divergence in DNA methylation for a given region in each epiHybrid, as a quantitative molecular trait and performed pairwise-correlation analysis. We uncovered striking co-variation patterns across the genome (Figure 3.12c). Most striking was the strong within- and between pericentromeric co-variation of NADs in non-CG context. We broke down the NAD correlations further and figured out that most of the correlated regions are SMR-TCdM events often located on other chromosomes (Figure 3.13). The extensive co-regulation of regions in *trans* suggests a possible *trans*-acting mechanism, perhaps mediated via sRNAs, therefore we further correlated the mid-parent divergence in 24nt sRNA abundance of the same regions and found a similar pattern (Figure 3.14).

These results highlight that methylome remodeling in epiHybrids is globally orchestrated and involves distally coordinated NAD events, with a substantial cross-talk among pericentromeric regions, particularly in non-CG context. These are at least partly the result of these regions being co-targeted by *trans*-acting sRNA.

3.4 Parental DMRs direct methylome remodeling locally and at distal regions

Since DNA methylation and sRNA remodelling events are occurring in so many independent epiHybrid families, indicated to us that they could be traced back to methylome features that are shared among the paternal parents. Previous work had shown that the epiRIL paternal parents segregate hypomethylated *ddm1*-induced regions across the genome, most prominently in pericentromeric regions (Figure 3.15). These segregating *ddm1*-induced hypomethylated regions

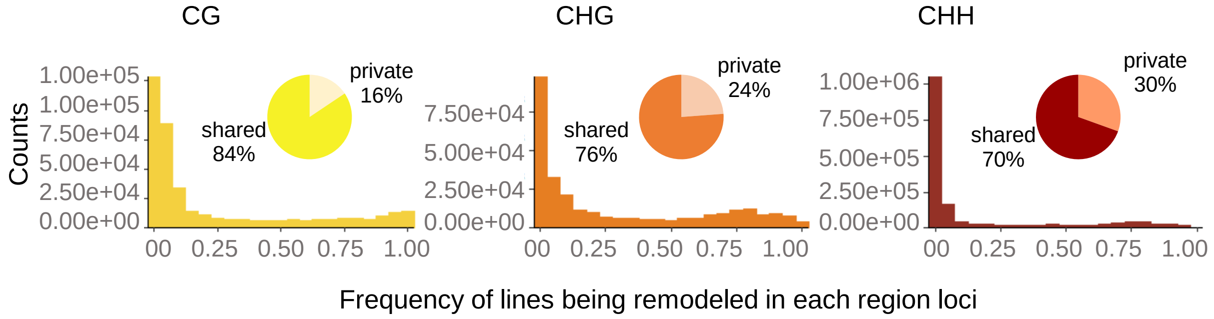


Figure 3.11: This figure displays the frequency distribution of each line for every 200 bp region. A frequency of 1 indicates that all 169 lines have remodeling regions for a specific 200 bp region (Kakoulidou et al. 2024).

are shared by about 25% of the epiRILs while the rest 75% are wild-type methylated (Cortijo et al. 2014; Colomé-Tatché et al. 2012; Zhang et al. 2021).

We made use of the epiRIL linkage map (Zhang et al. 2021) involving 144 markers in a QTL^{epi} mapping framework to explore if these segregating regions can be used to predict remodelling in the epiHybrids. We treated the degree of mid-parental methylation divergence at each NAD in the epiHybrids as a molecular quantitative trait and used the segregating parental DMRs as markers (predictors), thus performing one genome-wide linkage scan for each shared NAD. By design, this allowed us to assess if a given QTL^{epi} correlated with a specific NAD region locally or at distal regions (Figure 1.6). At each DMR marker, a given epiRIL parent can either be epihomozygous for the wild-type methylated state (MM) or epihomozygous for the *ddm1*-like state (UU) (Figure 3.15a). Clear DNA methylation patterns represents each marker's status, as the epiRILs are hypomethylated where there is a *ddm1* inherited haplotype, and hypermethyled where there is a wild-type haplotype (Figure 3.15b).

A significant proportion of regions, 15% for CG, 38% for CHG, and 16% for CHH, showed that NADs were associated with a QTLepi, as demonstrated in Figure 3.15. Moreover, these QTL effects were substantial, explaining on average 38% of the mid-parent methylation divergence in each NAD region (Figure 3.15c). Interestingly, the majority (94%) of all detected NAD-QTL^{epi} associations were located in distal regions and were mostly restricted to pericentromeric regions (88% of total). These associations were observed both intra-chromosomally (90%) and inter-chromosomally (10%) (Figure 3.17a), highlighting the importance of genetic regulation in the establishment of NADs.

We also categorized the NAD regions based on whether they were primarily associated with TCM or TCdM events across the epiHybrids. We observed that negative associations, which were mostly observed in distal regions (96%), were more likely to occur in epiHybrids with TCM events in pericentromeric NAD regions whose paternal epiRIL parents were methylated at theQTL^{epi}(Figure 3.17b). Likewise, TCdM events were more likely in epiHybrids whose paternal parents were unmethylated at the QTL^{epi}.

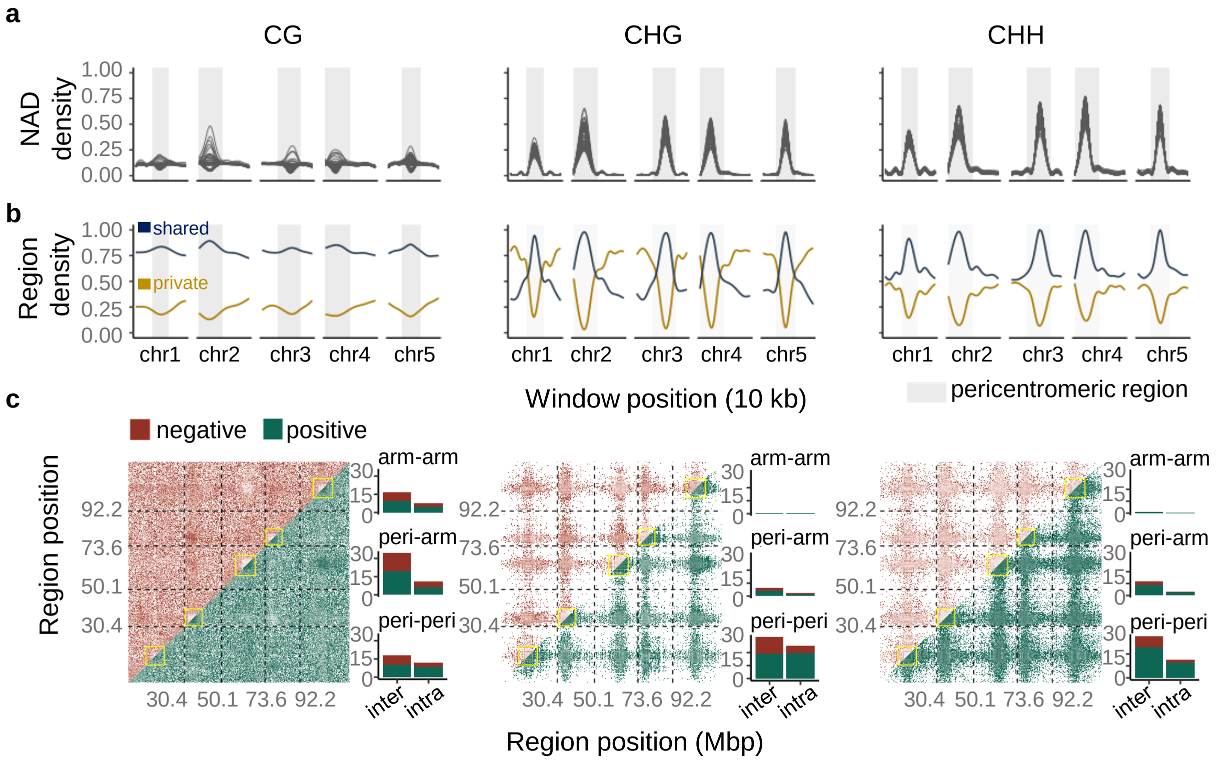


Figure 3.12: Pairwise-correlation analysis of remodeling regions. a) This figure displays the genome-wide density of NAD (non-additive DNA methylation) regions in 10 kb sliding windows with a step size of 10 kb. Each line represents an epiHybrid family, and grey areas indicate pericentromeric regions. b) Genome wide density of shared (blue) and private (yellow) regions in the same 10 kb sliding windows. c) Frequency density of significant pairwise correlations (p-value < 0.05) between NAD regions, shown in a genome-wide matrix with negative correlations in the upper triangle and positive correlations in the lower triangle. Only NADs shared by at least 10 epiHybrid families are considered, with color intensity indicating significance levels. The accompanying barplots show the frequency of significant correlations within and between chromosomes, categorized as 'arm-arm' when both NAD regions are in chromosome arms, 'arm-peri' when one region is in the arm and the other in the pericentromeric region, and 'peri-peri' when both regions are in the pericentromeric region. Grey areas indicate pericentromeric regions (Kakoulidou et al. 2024).

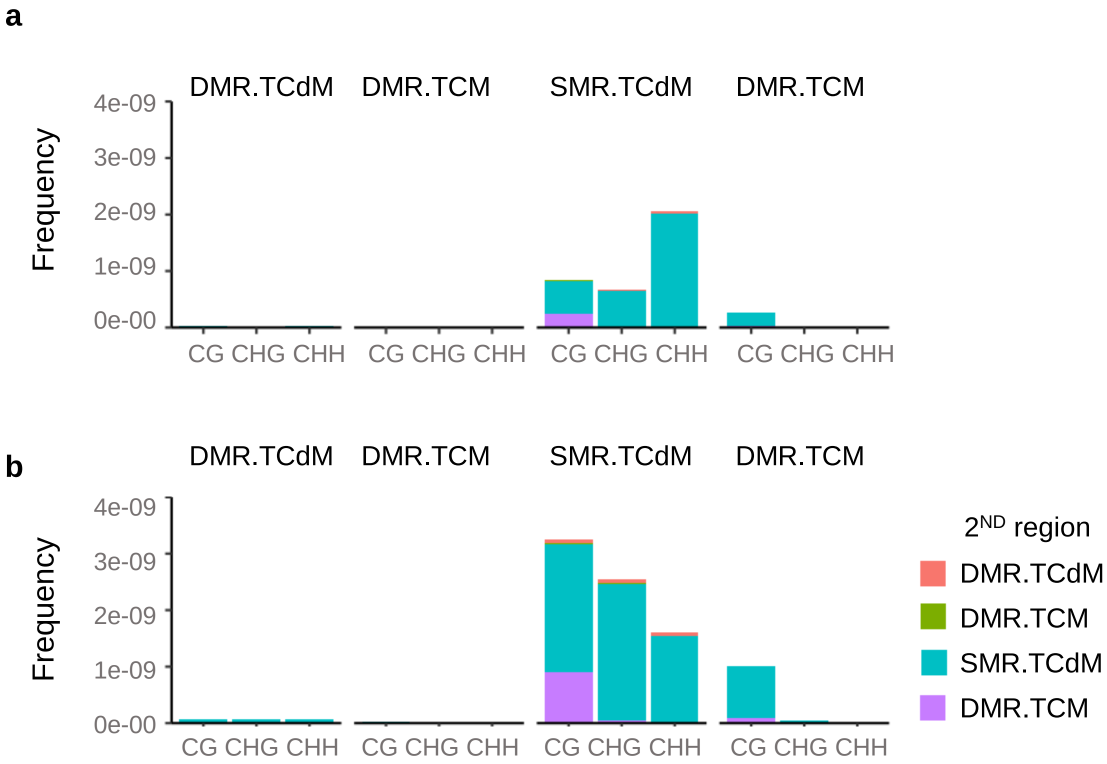


Figure 3.13: The graph shows the frequency of correlated regions, indicating whether both regions are located on the same chromosome (a) or on different chromosomes (b). The x-axis displays the identity of the first region of each pair, and the filled colors indicate the identity of the second correlated region (Kakoulidou et al. 2024).

In contrast, for the 245 positive associations, we observed that TCM events in NAD regions occurred more frequently in epiHybrids whose paternal parents were unmethylated at the QTL^{epi}, while TCdM events were more prevalent in epiHybrids whose paternal parents were methylated(Figure 3.17b). Interestingly, the NAD regions targeted by QTL^{epi} showed strong correlations with each other across the genome, suggesting that the QTL^{epi} exert pleiotropic effects that contribute to the genome-wide remodeling of distal regions in the epiHybrids."

It is noteworthy that each identified QTL^{epi} was linked with at least one remodeling region that exhibited strong correlations with each other in the epiHybrids. Therefore, the methylation status of specific loci in the paternal genome guides and can be utilized to predict how the epiHybrids' methylome remodeling is coordinated at a genome-wide level.

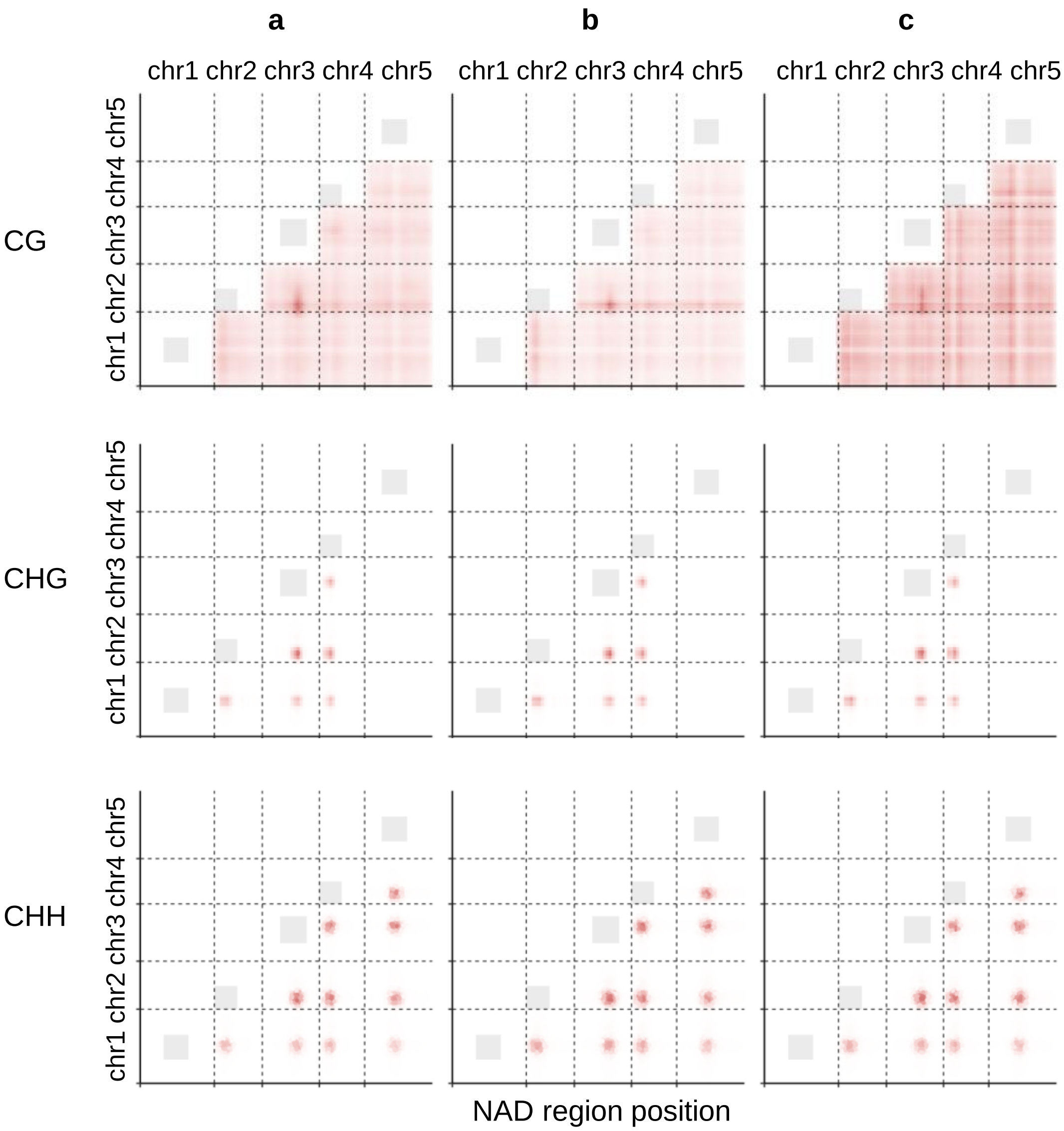


Figure 3.14: Frequency of significant correlations per context for selected inter-regions in sRNA and methylation (a), sRNA (b), and methylation (c) (Kakoulidou et al. 2024).

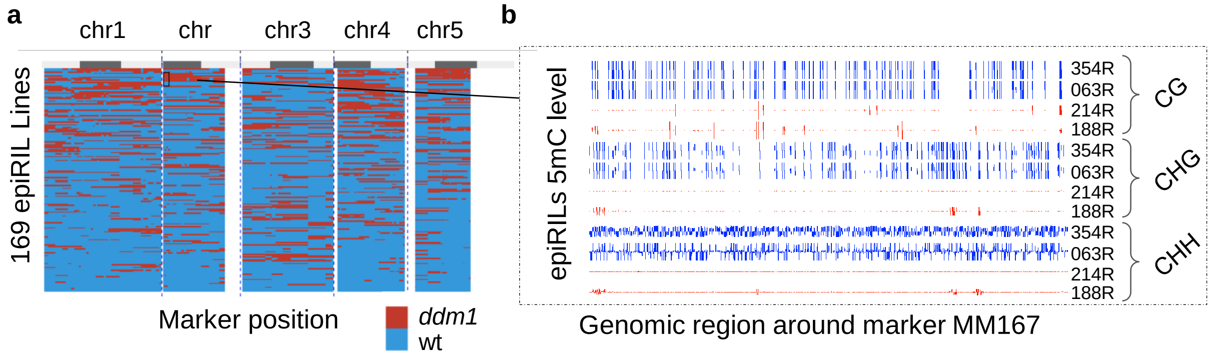


Figure 3.15: a) A haplotype map for the 169 epiRIL. Each row represents the genome composition of one epiRIL for the 144 DMRs. Red bars represent regions from the *ddm1* parent while blue regions from the wild-type parent. On the top, light gray bars indicate the chromosomes while dark gray bars indicate the pericentromeric regions. b) An example, genome browser shot (Robinson et al. 2011) for four epiRIL families at the marker MM167. Hypermethylated and hypomethylated epiRILs at these regions are indicated with blue and red color respectively (Kakoulidou et al. 2024).

3.5 Parental pericentromeric DMRs trigger heterosis

We observed that parental pericentromeric DMRs have a significant impact on DNA methylation remodeling in the epiHybrids, as shown in Figure 3.17a. This finding led us to hypothesize that these DMRs play a role in inducing phenotypic heterosis in the epiHybrids, potentially through their influence on the regulation of specific genes at the transcriptional level.

To analyze this further, we utilized an automated phenotyping facility (Junker et al. 2015) to phenotype 190 epiHybrid families. We conducted two phases of phenotyping, each consisting of three individual replication experiments. For our analysis of heterosis, we selected the second phase, as these plants were used for DNA sequencing. Images of small plants obtained from the IPK phenotyping facility (Junker et al. 2015) were processed using the open-source IAP (Integrated Analysis Platform) software developed by Klukas et al. (2014).

Heterosis was studied by comparing the phenotypic performance of the hybrids with that of their parents. We chose LA on the 18th day after sowing as our focal trait, as it has often been used as a measure of hybrid performance (Meyer et al. 2004). We define heterosis as the phenotypic divergence (in %) of an epiHybrid from the MPV, and we named it mid-parent heterosis (MPH). Among the epiHybrid families, there was substantial MPH, with some epiHybrids showing up to 31% increase and 26% decrease in LA, respectively (Figure 3.18a).

We used variance component analysis to estimate the proportion of total variation in MPH that could be attributed to between-family variation in epigenomic paternal states. Our findings showed that in the case of LA, the paternal methylome accounted for a significant portion (30%) of heterosis in the epiHybrids, highlighting its importance as a determinant of phenotypic variation. This observation is consistent with previous research conducted on a panel of 19 *ddm1*-epiHybrids,

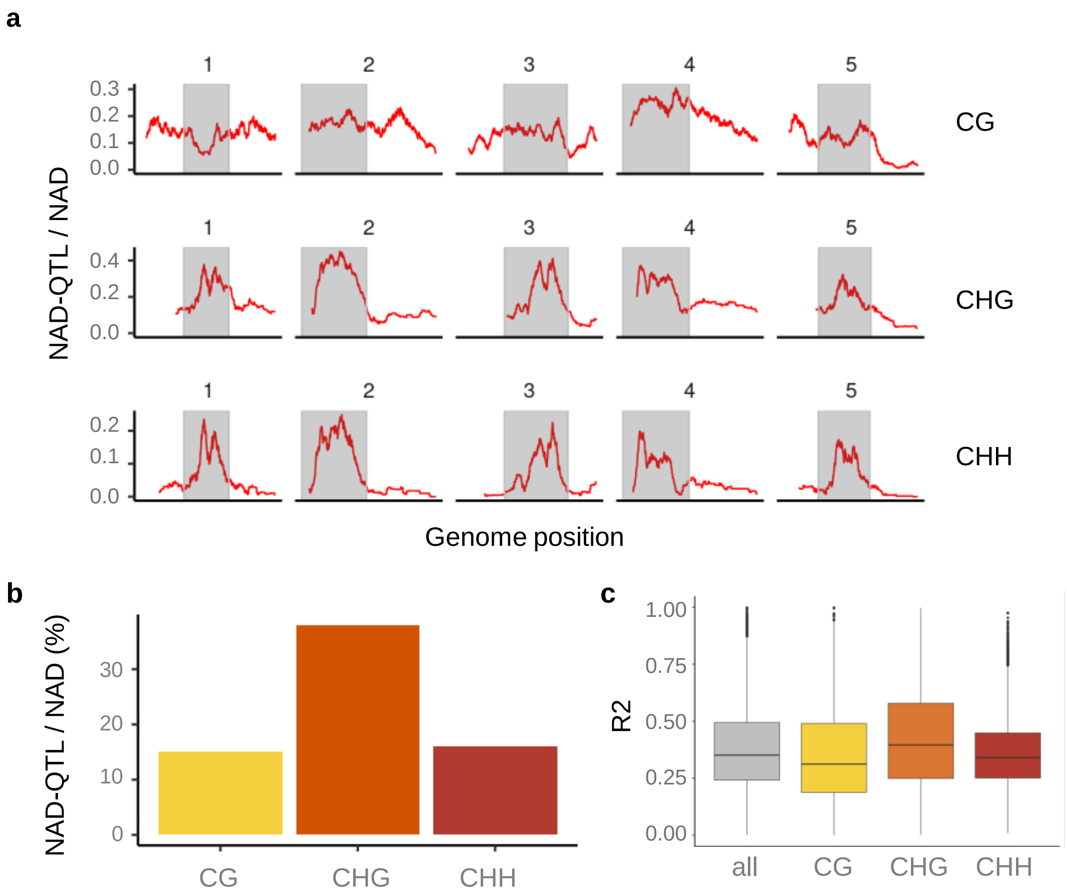


Figure 3.16: Significant QTL^{epi} regions. a) Frequency of significant NAD-QTL^{epi} targets used for the QTL analysis genome-wise. b) Average frequency of significant NAD-QTL^{epi} targets for each context. c) Coefficient of determination (R^2) for the linear regression model, where HD is the dependent variable and the epigenotype is the independent (Kakoulidou et al. 2024).

which also found a strong influence of paternal parents on phenotypic variation (Lauss et al. 2018).

To assess if the between-family variation in MPH can be explained by parental pericentromeric DMRs, we performed an unbiased genome-wide QTL^{epi} scan using parental DMRs as predictors and the degree of LA MPH within each epiHybrid family as the outcome variable (Figure 3.18 c). The analysis revealed that pericentromeric DMRs on chromosome 1, 2, and 3 were significantly associated with LA MPH (Figure 3.18C) and together explained 13% of the total between-family MPH variance (Figure 3.22d). Notably, our analysis revealed that for two chromosomes (chr2 and 3), there was no significant difference in heterosis between epiHybrids whose male epiRIL parent had DMRs that were unmethylated, resembling the *ddm1*-like state, and those whose paternal epiRILs had wild-type methylated DMRs (Figure 3.18d). In contrast, we observed an opposite effect on LA heterosis for chromosome 1.

As we tried to identify additional causative parental DMRs, we performed an unbiased genome-wide QTL^{epi} scan. Only three QTL^{epi} regions in total were identified, already in close LD with the candidate pericentromeric DMRs above. Hence, the pleiotropic NAD-QTL^{epi} on chr 1, 2 and 3 account for essentially all detectable parental methylome contributions to LA MPH in the epiHybrids (Kakoulidou et al. 2024).

A possible mechanism by which the pleiotropic QTL^{epi} affect phenotypic heterosis is by modifying the expression of the genes nearby to their remodeled target regions. Of the 806 NAD targets that were correlated with the peak LA QTL^{epi}, 594 (74%) were mapping within 1kb of 355 unique-protein coding genes. In order to test their association with expression, we performed RNA sequencing of the same 36 families that were used for the sRNA-seq analysis. Of the 594 NAD-QTL^{epi} associations, 84% (499) were still significant in the subset of 36 trios for the methylation divergence and epigenotype and mapped to 302 unique genes. Further filtering of the 499 NAD-QTL^{epi} revealed that 26 of the corresponding genes also had clear correlations in their degree of midparent divergence in expression and DNA methylation (Figure 3.19 and Table 3.1), while for 25 genes the corresponding NAD-QTL^{epi} also acted as epigenetic expression QTL (eQTL^{epi}) (Table 3.2).

Both filtering strategies gave some promising candidate genes. Among those, on the QTL^{epi} on chromosome 3, we identified a chromatin remodeling protein of the CLASSY family (CLSY4), DIHYDROXYACID DEHYDRATASE (DHAD) and CHLOROPLAST UNUSUAL POSITIONING 1 (CHUP1). CLSY4 controls the production of 24nt sRNAs that results in both locus-specific and global regulation of DNA methylation (Zhou et al. 2018). DHAD catalyses a key step in the branched-chain amino acid (BCAA) biosynthetic pathway and is involved in salt stress responses (Zhang et al. 2015). Knockout mutants result in a reduction in the accumulation of all three BCAAs in roots, but increase sensitivity to abiotic stressors (Zhang et al. 2015). On the other hand, CHUP1 is crucial for chloroplast movement in leaves in response to light (Oikawa et al. 2003). Without CHUP, chloroplasts are permanently clustered at the bottom of the palisade cells, and leaves are prone to photoinhibition. This results in damaged reproduction growth, as the plants produce fewer leaves and siliques and less leaf biomass (Howard et al. 2020). Interestingly, the QTL^{epi} on chr3 also mapped close to the QTL^{epi} identified by Lauss et al. (2018), which was associated with leaf area and flowering time in their smaller pilot study.

On chromosome 2, the target NAD-QTL is associated with the expression of a histone methyltransferase SU(VAR)3-9 RELATED 5 (SUVR5) (Figure 3.19a). SUVR5 is part of a multimeric complex that binds DNA through its zinc fingers and represses gene expression by altering histone modifications. Loss of function mutants display delayed flowering and reduced root growth (Caro et al. 2012). Interesting, Cortijo et al. (2014) also identified a QTL that controls root length in a distal location in chromosome 2. On chromosome 1, CARBAMOYL PHOSPHATE SYNTHETASE (CARB or VEN3) showed an increase in expression associated with the negative methylation divergence. VEN3 is involved in plant metabolism and responses to stress, while loss of function mutants show reticulated leaf phenotype and defective mesophyll development (Mollá-Morales et al. 2011).

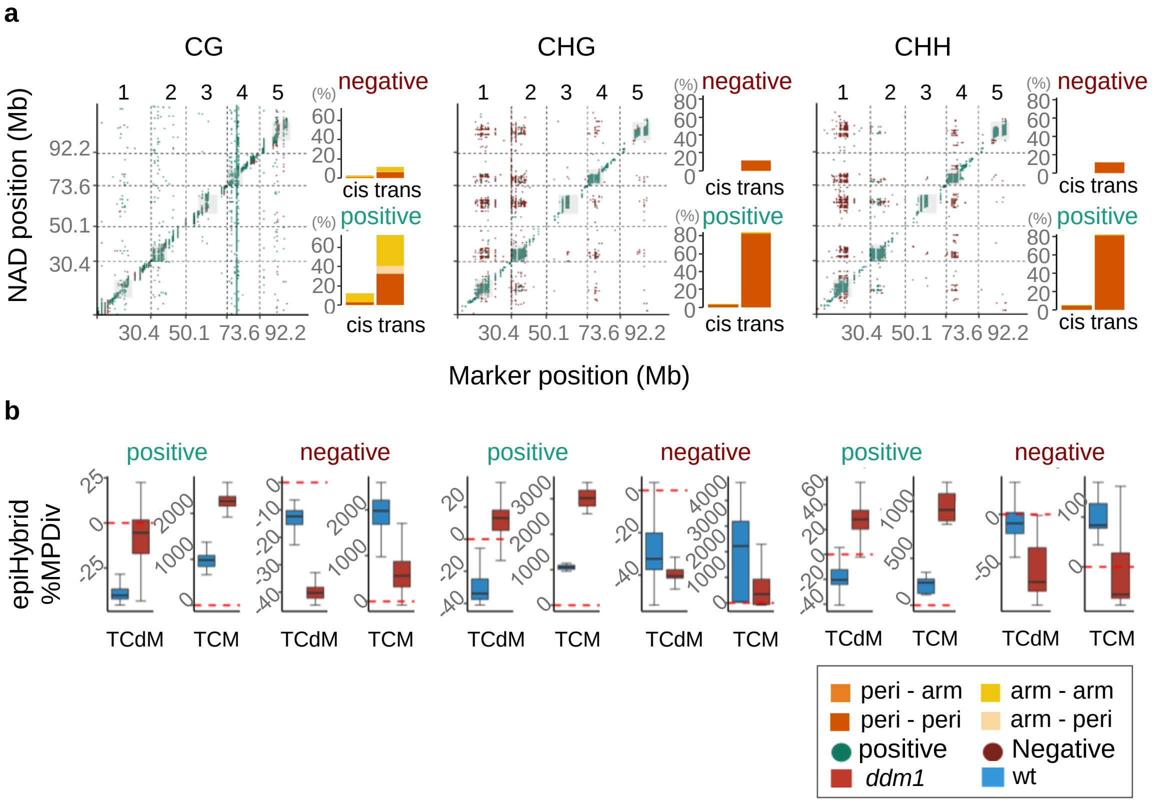


Figure 3.17: a) Genome-wide associations between DMR markers and NAD regions. Green dots indicate positive effect direction, while red dots show negative effect direction. The barplots on the side quantify the statistics of the significant NAD-QTL^{epi}s in the following categories; “arm - arm” means that both the marker and the NAD target region are in the chromosome arm; “arm - peri” indicates that the marker is located in the chromosome arms but the NAD - target region in the pericentromeric region; “peri - arm” means that the marker is located in the pericentromeric but the NAD-target region at the chromosome arm; “peri - peri” that both the marker and the NAD-target region are located in the pericentromeric region. b) The boxplots show the percent of midparent divergence in DNA methylation in the epiHybrids at the NAD targets, categorized by QTL^{epi} effect direction (positive or negative) and predominant NAD remodeling scenario (TCdM or TCM). Blue: midparent divergence in epiHybrids whose paternal epiRILs were epihomozygous wildtype (MM) at the QTL^{epi}; red: midparent divergence in epiHybrids whose paternal epiRILs were epihomozygous *ddm1* (UU) at the QTL^{epi}.(Kakoulidou et al. 2024).

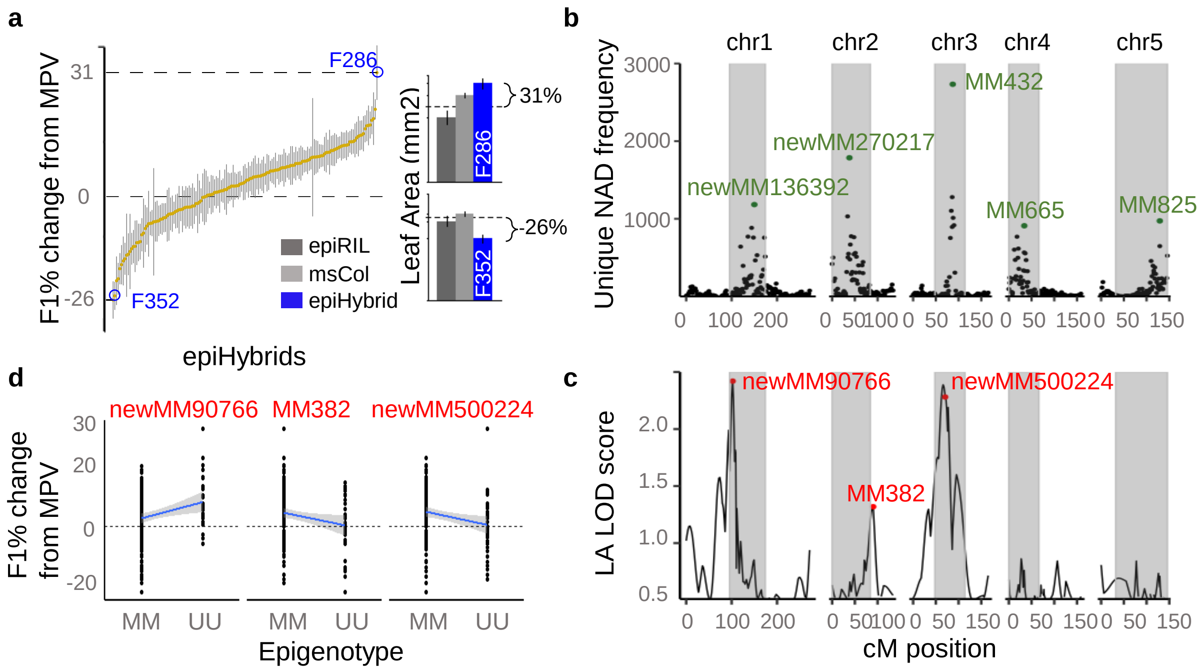


Figure 3.18: LA distribution for each epiHybrid family with 2 extreme examples shown in the barplots. b) Average LOD score profile for NAD regions uniquely associated with a given QTL^{epi}. Marked with green are the most pleiotropic DMR markers. c) A QTL^{epi} scan for LA heterosis reveals 3 QTL regions in pericentromeric regions on chromosomes 1, 2 and 3. d) Effect size and direction of the detected QTL^{epi} (MM epihomozygous wt, UU epihomozygous *ddm1*) for LA.

To obtain first mechanistic insights into how the pleiotropic LA heterosis QTL^{epi} affect these candidate genes, we performed causal modelling. We found that for most genes (58%) the pleiotropic QTL^{epi} affect DNA methylation and expression independently. Nonetheless, consistent with Meng et al. (2016), a substantial proportion (42%) of QTL^{epi} also affect gene expression indirectly via effects on DNA methylation at proximal NADs (Figure 3.20), rather than the latter occurring in an expression-dependent manner ((Secco et al. 2015)). However, 90% of these latter NAD-QTL^{epi} associations were in at distal regions, with the majority of the NAD targets being located within 2Mb outside of the QTL^{epi} confidence interval (Figure 3.20).

These distal associations therefore require some type of long-range signal by which the QTL^{epi} can affect the NAD status. One possibility is that the differential production of distal-acting sRNA from DMRs within the QTL^{epi} confidence interval leads to differential targeting of the NAD regions. Preliminary support for this comes from the fact that variation in sRNAs among epiHybrids for 13% of the same NAD-QTL^{epi} associations correlate with the QTL^{epi}-induced methylation variation of the NAD target regions. Follow-up molecular work is required to further delineate a sRNA-based mechanistic model underlying these at distal regions effects.

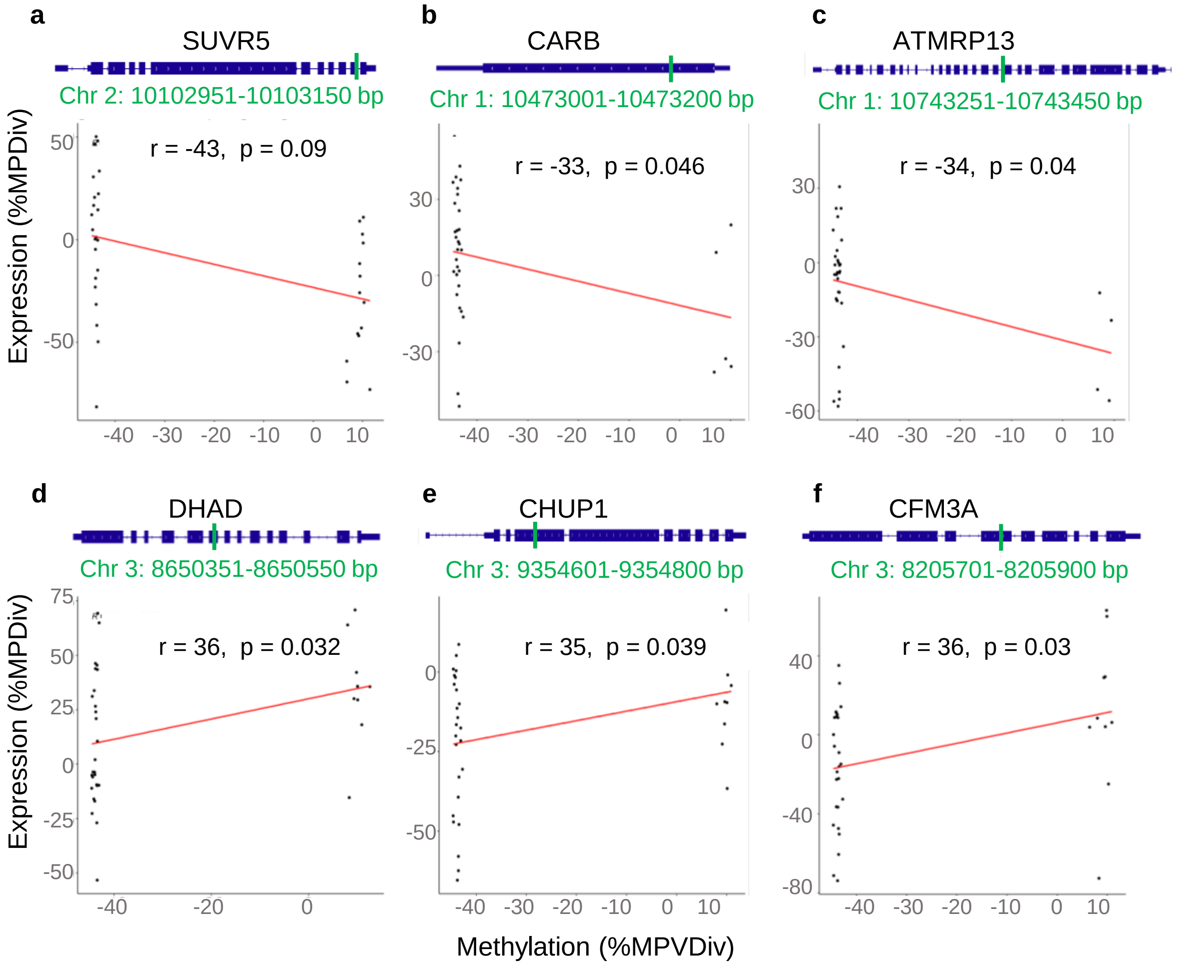


Figure 3.19: Changes in DNA methylation levels in the epiHybrids are associated with corresponding changes in gene expression levels. The top of the figure displays a schematic representation of each gene, with the green bar indicating the corresponding NAD-QTL^{epi} target region (Kakoulidou et al. 2024).

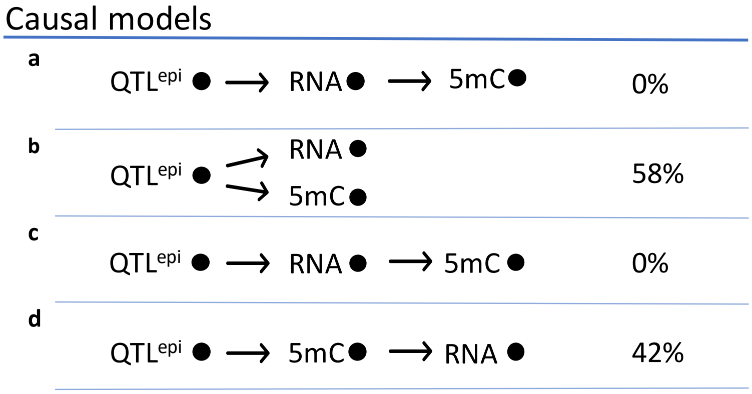

Causal models
a
QTLepi ● ⟶ RNA● ⟶ 5mC●
0%
b
QTLepi ●
RNA ●
5mC●
58%
c
QTLepi ● ⟶ RNA ● ⟶ 5mC ●
0%
d
QTLepi ● ⟶ 5mC● ⟶ RNA ●
42%

Figure 3.20: Model “a” indicates that there is no direction in the relationship between gene expression, methylation and QTL^{epi}. Model “b” shows that QTL^{epi} act on methylation and gene expression independently. Model “c” shows how QTL^{epi} act on methylation through gene expression and model “d” shows how QTL^{epi} acts on gene expression through methylation (Kakoulidou et al. 2024).

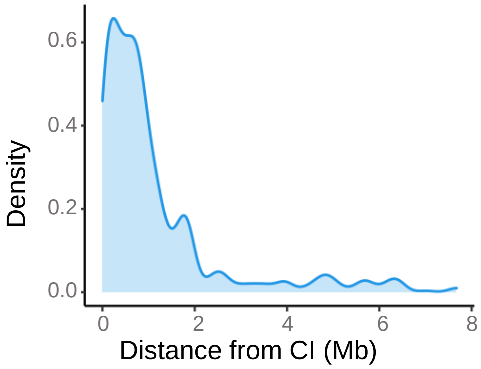


Figure 3.21: Distance between the NAD-QTL targets in distal regions and their corresponding confidence interval boundaries (Kakoulidou et al. 2024).

3.6 Phenotypically relevant *de novo* NADs

Our examination of the methylomes of 169 epiHybrids identified regions containing *de novo* NADs. These NADs could not be easily attributed to either locally or distal-acting parental QTL^{epi}, and were found to be shared among families. Based on this, we hypothesized that the *de novo* NADs might have originated in the paternal methylomes and could serve as significant predictors of phenotypic heterosis. To investigate whether DNA methylation changes at the *de novo* NADs were associated with phenotypic variation in LA, we conducted an EWAS study that controlled for the effects of previously identified parental QTL^{epi} on chromosomes 1, 2, and 3.

Our EWAS identified a large number of significant associations (Figure 3.22a), which accounted for approximately 30% of the between-family variation in LA MPH (Figure 3.22d). For most of the *de novo*, an increase in DNA methylation in the epiHybrids showed a positive correlation with heterosis (Figure 3.22c). The combination of parental QTL^{epi} and de novo NADs in the epiHybrids appear to be an important molecular component underlying heterosis, as they can explain up to 51% of the between-family variation in LA MPH in the epiHybrids (Figure 3.22d). Although the remaining sources of variation in our experimental design remain unknown, there is currently no evidence that the genetic variants recently detected in the epiRILs and in siblings of their *ddm1-2* founder line contribute (Figure 3.24).

The contribution of similar parental DMRs to heterosis in classical hybrid crosses, where the two parents are genetically distinct, is not well understood. Nevertheless, we hypothesise that many heterosis QTL that have been traditionally attributed to parental genetic polymorphisms, might be the result of being in close LD with segregating hypo-methylated epialleles that trigger similar methylome remodelling dynamics. To support this hypothesis, we examined known heterosis QTL identified in classical Arabidopsis hybrid studies that explain the same trait as in our investigation. We observed that several of these QTL are located in proximity to our detected QTL^{epi} on chromosomes 1, 2, and 3, providing some evidence for this idea (Figure 3.23).

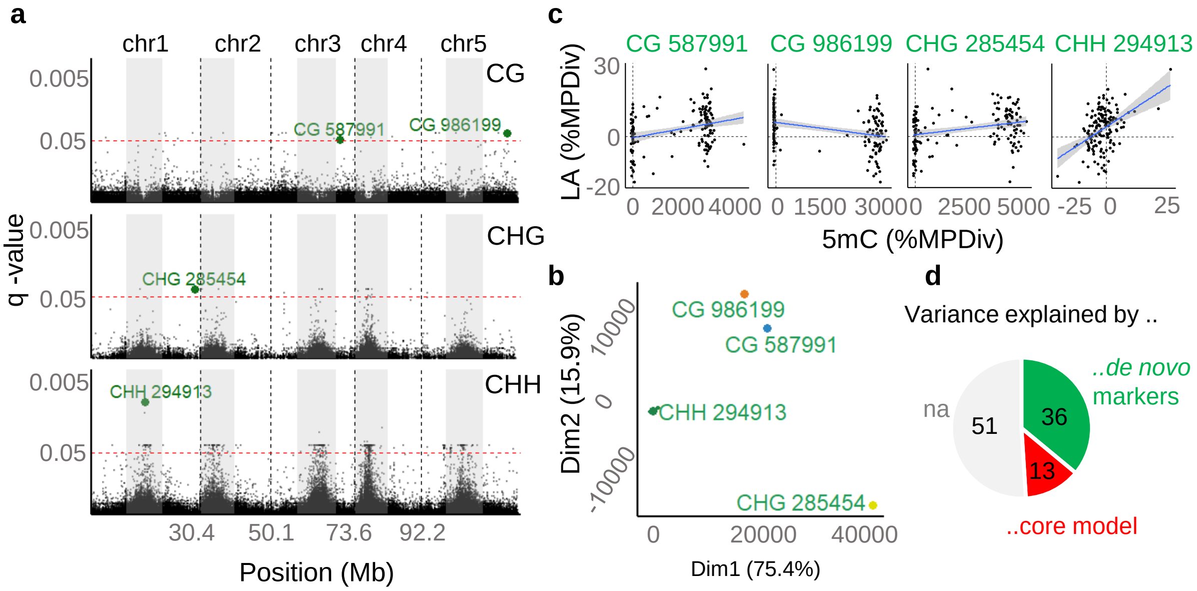


Figure 3.22: Epigenome-wide association study. a) The Manhattan plot indicates the significant associations of the EWAS scan. *P*-values were adjusted using Benajmini-Yekutieli multiple comparisons and are plotted on the y-scale (q-value). The most significant markers are marked with green. b) Results of the Hierarchical Clustering on Principal Components (HCPCA) (q-value < 0.05). c) Correlations between epiHybrid phenotypic and methylation divergence at the selected *de novo* markers. d) Results of variance component analysis used to estimate how much of the phenotypic divergence of LA can be attributed to parental DMR markers (core model) or to *de novo* markers(Kakoulidou et al. 2024).

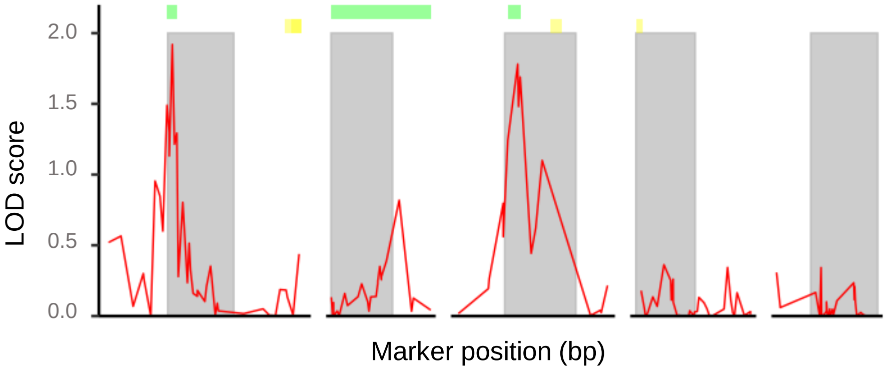


Figure 3.23: Genome-wide QTL scan for LA. Marked with green are the confidence intervals of our significant peaks while with yellow are marked, published QTLs for mid-parental heterosis in leaf area (Kakoulidou et al. 2024).

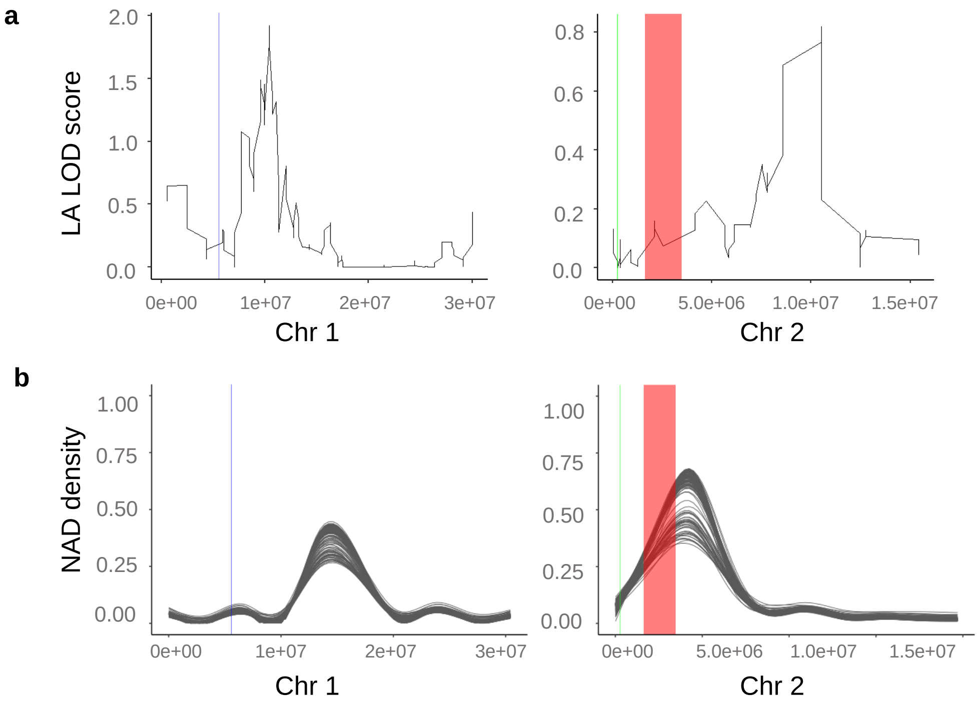


Figure 3.24: Red bar indicates the detected Chr2-2M inversion. Green line indicates the 56kb inversion and blue indicates the 55kb inversion, as described in Zhang et al., 2022 (under review). They are marked in the (a) LA LOD score profile figure (Figure 3.18) and in the (b) remodeling figure (Figure 3.12) (Kakoulidou et al. 2024).

Table 3.1: Genes where methylation divergence correlates with expression divergence.

IDs	Symbol	Chr	Position (Mbp)	Description
AT1G28010	ABCB14	1	9.763211	ABC transporter B family member 14
AT1G28350		1	9.943737	Tyrosine–tRNA ligase 2, cytoplasmic
AT1G29240		1	10.216779	F28N24.8 protein
AT1G29320		1	10.25525	Transducin repeat-like superfamily protein
AT1G29900	CARB	1	10.467956	Carbamoyl-phosphate synthase chain
AT1G30300		1	10.672973	At1g30300
AT1G30410	ATMRP13	1	10.738683	multidrug resistance-ass protein
AT1G30755		1	10.905731	Elongation factor G, putative (DUF668)
AT2G23740	SUVR5	2	10.097265	Histone-lysine N-methyltransferase
AT2G24220	ATPUP5	2	10.300032	purine permease 5
AT2G24340		2	10.35543	Sequence-spec. DNA binding transcr. factor
AT2G24600		2	10.452034	Ankyrin repeat family protein
AT2G25140	CLPB4	2	10.697629	Chaperone protein ClpB4
AT2G26270		2	11.183842	BRCT domain-containing DNA repair protein
AT2G26280	CID7	2	11.187726	protein-interacting protein 7
AT2G26510	NAT3	2	11.273842	Nucleobase-ascorbate transporter 3
AT2G26810		2	11.433662	Putative methyltransferase family protein
AT3G22260		3	7.870826	At3g22260
AT3G22980		3	8.160075	Ribosomal protein S5
AT3G23070	CFM3A	3	8.203393	CRM-domain containing factor CFM3A
AT3G23080		3	8.207287	Membrane related protein-like
AT3G23940	DHAD	3	8.648669	Dihydroxy-acid dehydratase, chloroplastic
AT3G25690	CHUP1	3	9.352444	Hydroxyproline-rich family protein
AT3G26290	CYP71B26	3	9.632608	Cytochrome P450,family 71
AT3G28050		3	10.442608	WAT1-related protein
AT3G45050		3	16.475957	AT3G45050 protein

Table 3.2: Genes where the NAD-QTL epigenotypes correlates with expression divergence.

IDs	Symbol	Chr	Position (Mbp)	Description
AT1G28110	SCPL45	1	9.803641	Serine carboxypeptidase-like 45
AT1G30300		1	10.672973	At1g30300
AT1G30410	ATMRP13	1	10.738683	multidrug resistance-ass. protein
AT1G30530	UGT78D1	1	10.814612	Glycosyltransferase
AT2G24330		2	10.347018	Uncharacterized protein At2g24330
AT2G24340		2	10.35543	Sequence-spec, DNA binding transcr. factor
AT2G24600		2	10.452034	Ankyrin repeat family protein
AT2G25297		2	10.770676	
AT2G26990	CSN2	2	11.519388	COP9 signalosome complex subunit 2
AT3G22260		3	7.870826	At3g22260
AT3G22520		3	7.97472	Spindle assembly abnormal protei
AT3G23020		3	8.177002	Pentatricopeptide protein At3g23020
AT3G23070	CFM3A	3	8.203393	CRM-domain containing factor CFM3A
AT3G23340	CKL10	3	8.350803	Casein kinase 1-like protein 10
AT3G23490	CYN	3	8.422649	Cyanate hydratase
AT3G23510		3	8.427557	Cyclopropane-fatty-acyl-phosph synthase
AT3G23940	DHAD	3	8.648669	Dihydroxy-acid dehydratase, chloroplastic
AT3G24030	THIM	3	8.679819	Hydroxyethylthiazole kinase
AT3G24340	CLSY4	3	8.831935	SNF2 domain-containing protein CLASSY 4
AT3G24515	UBC37	3	8.934408	Ubiquitin-conjugating enzyme 37
AT3G25970	PCMP-E46	3	9.500016	
AT3G26330	CYP71B37	3	9.646784	Cytochrome P450 71B37
AT3G26810	AFB2	3	9.867652	AFB2
AT3G27730	RCK	3	10.273098	DNA helicase ROCK-N-ROLLERS
AT3G28050		3	10.442608	WAT1-related protein

Chapter 4

Discussion

Heterosis, also known as hybrid vigor, is a fascinating phenomenon observed in the F1 hybrid offspring that displays an improved phenotype compared to its parents. The concept of heterosis has been studied for decades in agricultural breeding and has played a crucial role in the development of high-yielding crops. The increased crop performance achieved through hybridization has revolutionized modern agriculture and has been instrumental in addressing global food security challenges.

The genetic approaches aimed to understand the mechanisms underlying heterosis have highlighted the importance of the interaction between different parental DNA sequence alleles. The hybrid vigor observed in the F1 hybrid offspring is believed to result from the complementation of deleterious recessive alleles that have become fixed in each of the parental varieties and are coming together in the hybrid.

However, genotypic information is often not enough to successfully predict heterosis. There is growing evidence that beside genetic also epigenetic factors contribute to heterosis and that heterotic hybrids display widespread changes in DNA methylation and sRNAs relative to their parental lines (Lauss et al. 2018; Zhang et al. 2016; Groszmann et al. 2011a). The extent to which this remodeling can be attributed to specific DMRs in the parental genomes and the contribution of the DMRs to phenotypic heterosis remain poorly understood (Lauss et al. 2018).

Here we used near-isogenic but epigenetically divergent parents, to create the so-called epigenetic Hybrids (epiHybrids). This experimental system allowed us opportunity to test whether epigenetic divergences between the parental lines are associated with heterosis in F1 hybrids independently of genetic differences. Phenotypic analysis in these epiHybrids revealed strong heterotic phenotypes for leaf area, indicating that epigenetic divergence between the parental lines is sufficient to cause heterosis, independently of DNA sequence determinants.

In this experiment, 169 epiRILs were used as the paternal parents and their high-resolution methylome sequences are an important by-product of this study. The epiRILs have been used in several experimental phenotyping and molecular studies (Lauss et al. 2018; Roux et al. 2011; Cortijo et al. 2014; Latzel et al. 2013; Kooke et al. 2015; Marí-Ordóñez et al. 2013) and continue to be a useful resource for plant epigenetic research. Molecular profiling of the epiHybrids now, showed that their epigenomes are substantially remodelled with respect to their parental lines, leading to epigenetic states that diverge both positively and negatively from the expected MPV.

Remodeling occurs both in DMRs and in SMRs, with the later being more frequent, as reported elsewhere (Zhang et al. 2016; Lauss et al. 2018; Ma et al. 2021). The remodeling events in these regions, referred to as TCM or TCdM, are mainly focused on clustered cytosines within DMRs or SMRs. Our pipeline identifies methylome remodeling regions as chromosome regions of at least 200 bp, which is consistent with the idea that heritable changes in the methylation status of cytosine clusters are functionally more relevant than changes in individual cytosines (Cortijo et al. 2014; Zhang et al. 2016).

DMR-associated remodelling events like TCM involves 24nt small RNAs that are initially produced from the methylated parental allele, but are subsequently targeting the unmethylated allele for *de novo* methylation (Greaves et al. 2012a; Shivaprasad et al. 2012; A-L et al. 2019; Chandler 2007). Indeed, TCM-DMRs are accompanied by a local increase in 24nt sRNA abundance in the epiHybrids, with levels equaling or exceeding those of the homozygous methylated parent. Interestingly, these sRNA differences can strongly predict heterosis in hybrid offspring, as it was shown in maize Seifert et al. (2018a), which is consistent with the pericentromeric origin of these effects. A useful follow-up study to verify the involvement of the RdDM pathway in methylome remodeling would be to use mutant plants deficient in RdDM activity, such as the *pol IV* or *ddm1/pol4* double mutant, and examine whether methylome remodeling in the epiHybrids is inhibited.

One pattern we observed is when parental regions are differentially methylated, but the hybrid loses methylation compared to the expected MPV (TCdM-DMR). Previous research suggests that these events occur more frequently in regions with greater genetic parental variation, indicating that RdDM of one allele requires the involvement of the other allele. According to the authors of that study, sRNA levels from the high-parent allele may not be able to induce DNA methylation in the low-parent allele because they are too divergent, resulting in a loss of methylation in the high-parent allele in the hybrids (Zhang et al. 2016). However, since we also observed DMR-loss events in our isogenic system at nearly the same frequency, this model is difficult to explain. As there is no clear ”trans-chromosomal” signal locally, we hypothesize that the trigger may originate from remodeling events in distal regions.

Mechanistic models explaining DMR-based remodeling events are built on molecular interactions between parental alleles, but such models cannot be easily extended to SMRs, where both parental alleles have the same methylation state. Therefore, the initiation of remodeling events in SMRs must come from ”distal” regions outside of homologous alleles. Here we propose a ”TCM proximity model” (Figure 4.1) , which suggests that many SMR gain events occur nearby TCM sites, indicating that they are a byproduct of a TCM process in the genomic ”neighborhood”. Although this model’s many details remain untested, one could hypothesize that sRNA-mediated targeting of the unmethylated allele at a DMR leads to imprecise DNA methylation spreading into flanking regions on the targeted allele, a process well-documented in the context of TE biology (Hollister and Gaut 2009; Diez et al. 2014).

Whether this spreading is monoallelic or not could not be directly verified in the isogenic hybrid system. However, many SMR gain events flanking TCM loci display clear intermediate methylation levels, which is at least consistent with this idea. Furthermore, many SMR gain events show a correlative increase in sRNA abundance in the flanking regions relative to the parents, providing additional support for the TCM proximity model. It would be interesting to

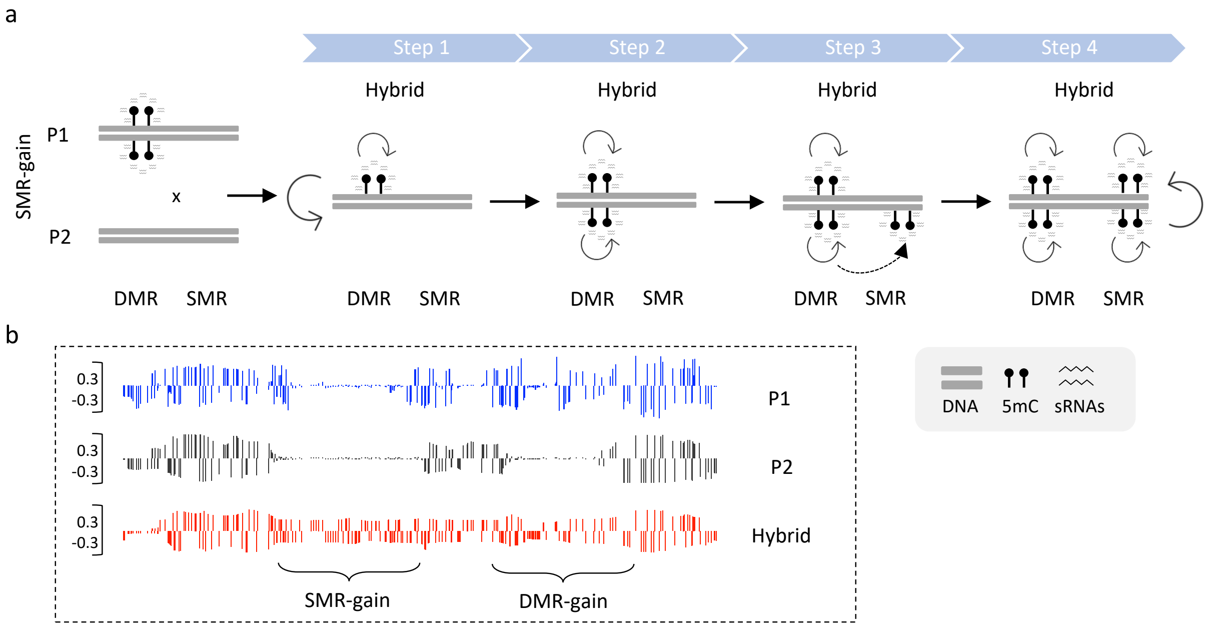


Figure 4.1: A TCM-proximity model for SMR-gain events. a) SMR-gain events are the result of DNA methylation spreading from TCM at proximal DMRs (Steps 1-3). Spreading is a monoallelic processes (Step 3). It is plausible that spreading leads to a second round of TCM in the opposite direction, thus producing biallelic gains of methylation at the SMR (Step 4). b) Example browser shots for CHH methylation context (Kakoulidou et al. 2024).

investigate whether the spreading leads to occasional re-targeting of the original donor allele for a second round of TCM, but in the opposite direction, explaining why many SMR gain events appear to be biallelic. A comprehensive evaluation of the TCM proximity model would necessitate allele-specific analyses, such as genetically diverse crosses, along with time-resolved data covering different developmental stages.

It has been proposed that the proximity model, as described above, could potentially explain SMR loss events as a byproduct of TCdM at DMR loss sites. However, since the molecular mechanism that triggers DMR loss is not yet known, such a model would not provide a satisfactory explanation. Additionally, SMR loss events are often located much further away from DMR loss events compared to SMR gain events making the proximity argument invalid.

Alternatively, it has been hypothesized that SMR loss events may be a result of remodeling of distal regions (Figure 4.2). This hypothesis suggests that SMR loss at a specific genomic locus may be correlated with methylation remodeling elsewhere in the genome. Our study provided a unique opportunity to investigate this hypothesis. Unlike previous studies that mainly focused on individual genomic regions, our analysis revealed that remodeling events co-vary substantially among distal regions both within and across chromosomes. This suggests that there is a significant cross-talk among pericentromeric regions, particularly in the non-CG context.

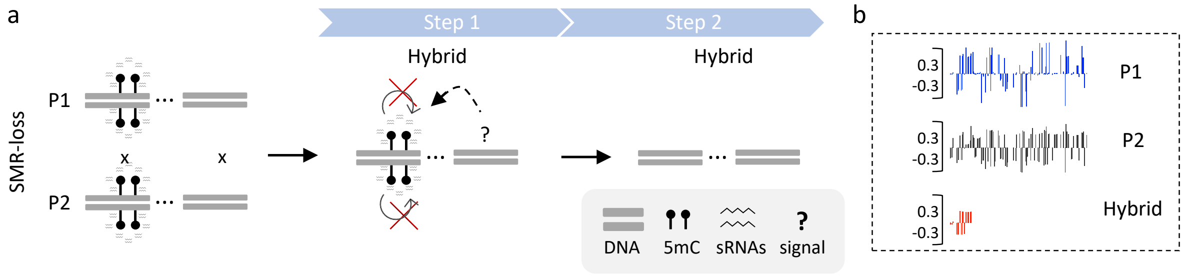


Figure 4.2: A “distal” model for SMR-loss events. a) SMR-loss events cannot be explained by TCdM of neighboring DMRs as they do not tend to occur in close proximity. Instead, current data indicate that SMR-loss events are trigger by a distal signal, possibly originating from a DMR remodeling event elsewhere in the genome. The molecular basis of the distal signal is currently unknown. b) Example browser shots for CHH methylation context (Kakoulidou et al. 2024)

For example, DNA methylation loss in one region was associated with methylation loss in another region in the hybrids, while gain in one region was associated with loss in another. Notably, pericentromeric regions of chromosomes emerged as major hubs, displaying high correlation not only locally but also between pericentromeric regions of different chromosomes and with euchromatic loci, including genes.

While the precise regulatory mechanisms underlying this cross-talk are not yet clear, our data suggest that trans-acting sRNA may play a crucial role in this process.The fact that we observed substantial co-variation of remodeling events among distal regions is particularly noteworthy. This suggests that there may be complex regulatory mechanisms at work that coordinate the remodeling of DNA methylation patterns across multiple regions of the genome.

Distally correlated remodeling events may be the result of homeostatic adjustments in hybrid DNA methylomes, which occur when new parental epigenomes are combined. This balancing act is achieved by methylases and demethylases, which regulate global or locus-specific methylation levels toward specific target values (Williams and Gehring 2020). An example of such regulation is the autoregulation of the demethylase REPRESSOR OF SILENCING 1 (ROS1), where the ROS1 promoter acts as a methylation sensor. Hypermethylation of the promoter sequence, via the RdDM pathway, results in the upregulation of ROS1, which initiates demethylation in a self-regulatory negative feedback loop, ultimately leading to decreased ROS1 expression (Williams and Gehring 2020). It is plausible that similar homeostatic mechanisms underlie the correlated DNA methylome remodeling patterns found in F1 hybrids, though quantifying these processes presents a challenge, as the system’s target values and all relevant ’sensor’ regions are not fully known (Kakoulidou et al. 2024).

Since DNA methylation and sRNA remodelling events are recurrent in so many independent epiHybrid families made us hypothesize that they can be traced back to methylome characteristics that are shared among the paternal parents. To explore this we used parental DMRs present in the epiRIL paternal lines in an QTL^{epi} mapping framework (Cortijo et al. 2014) to yield associations both locally and in distal regions. Parental DMRs were used as predictors and the level of remodeling at any given region as the molecular phenotype. Importantly, the 169 epiRIL methylomes from the present study were employed to improve the existing linkage map (Zhang et al. 2021) and the re-calibrated linkage map is an essential component of the present project. Firstly, it allowed us to identify heritably DMRs between the selected epiRILs and the Col-wt parental line. Second, these DMRs were used to perform an QTL^{epi} to demonstrate how these parental DMRs regulate methylation changes in the hybrids not only locally by also in distal regions.

In the epiHybrids' methylomes, we also found loci that showed *de novo* remodeling regions whose origin could not be easily attributed to local or distal-acting parental DMRs. However, these *de novo* NADs are shared among families and may therefore have a common origin in the paternal epiRILs methylomes. To examine if these *de novo* NADs is associated with the phenotypic performance of the epiHybrids we did a conditional EWAS, but taking into account the effects that were already controlled by the parental pleiotropic markers on the QTL^{epi} on chr 1, 2 and 3. Variance component analysis showed that the *de novo* NADs could account for 36% of the between-family variation in LA MPH and in combination with the parental QTL^{epi}, explain a major fraction (51%), indicating that they are an important molecular component underlying heterosis.

By analyzing gene expression in the same plant material that was used for WGBS, we looked into correlation between changes in gene expression and DNA methylation. Our findings revealed that changes in methylation, induced by *trans*-factors, have a significant impact on the expression of numerous target genes, which together contribute to phenotypic heterosis. Notably, many of the candidate genes that we identified were enriched in pathways related to chromatin function.

By utilizing integrative QTL mapping approaches to analyze phenome, methylome, and transcriptome data from a large panel of epiHybrid families, we were able to investigate the causal relationship between changes in DNA methylation and transcript levels at both local and distal regions, and their impact on hybrid performance. While genetic effects can account for changes in both DNA methylation and expression in most genes independently, our findings strongly suggest that effects on gene expression may also act via a direct effect on DNA methylation in proximal remodeling regions. Understanding how these methylome interactions respond to phenotypic heterosis could lead to the treatment of hybrid methylation as the output of a dynamic system, making it a useful predictor of heterosis.

Overall, the causal role of DNA methylation changes in driving non-additive gene expression and phenotypic heterosis in F1 hybrids is not fully understood. Current mechanistic models are limited to genomic regions where parental lines differ in DNA methylation status, while most remodeling events occur in regions with similar methylation patterns. The molecular basis of these events is unknown, but recent data suggest a major involvement of distally acting factors. Despite gaps in our understanding, evidence links parental DNA methylation and sRNA abundance with heterosis, which may serve as biomarkers in crop breeding.

List of Publications

Kakoulidou, I., Piecyk, R. S., Meyer, R. C., Kuhlmann, M., Altmann, T., Gutjahr, C., & Johannes, F. (2024). Mapping parental DMRs predictive of local and distal methylome remodeling in epigenetic F1 hybrids. Life science alliance, 7(4)., doi: 10.26508/lsa.202402599

Kakoulidou, I., & Johannes, F. (2023). DNA methylation remodeling in F1 hybrids. The Plant Journal. doi: https://doi.org/10.1111/tpj.16137

Hüther, P., Hagmann, J., Nunn, A., **Kakoulidou, I.**, Pisupati, R., Langenberger, D., ... & Becker, C. (2022). MethylScore, a pipeline for accurate and context-aware identification of differentially methylated regions from population-scale plant whole-genome bisulfite sequencing data. Quantitative Plant Biology, 3, e19. doi: https://doi.org/10.1017/qpb.2022.14

Kakoulidou, I., Avramidou, E. V., Baránek, M., Brunel-Muguet, S., Farrona, S., Johannes, F., ... & Maury, S. (2021). Epigenetics for crop improvement in times of global change. Biology, 10(8), 766. doi: https://doi.org/10.1017/qpb.2022.14

Zhang, Y., Jang, H., Xiao, R., **Kakoulidou, I.**, Piecyk, R. S., Johannes, F., & Schmitz, R. J. (2021). Heterochromatin is a quantitative trait associated with spontaneous epiallele formation. Nature Communications, 12(1), 6958. doi: 10.1038/s41467-021-27320-6.

Bibliography

A-L, L. G., C, L.-P., A, D., and S, M. (2019). Developmental, genetic and environmental variations of global DNA methylation in the first leaves emerging from the shoot apical meristem in poplar trees. *Plant Signaling & Behavior*, 14(6):1596717.

Andrews, S. a. o. (2010). FastQC: a quality control tool for high throughput sequence data. *Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom*.

Auwera, G. A. V. d. and O'Connor, B. D. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. "O'Reilly Media, Inc.". Google-Books-ID: vsXaDwAAQBAJ.

Barber, W. T., Zhang, W., Win, H., Varala, K. K., Dorweiler, J. E., Hudson, M. E., and Moose, S. P. (2012). Repeat associated small RNAs vary among parents and following hybridization in maize. *Proceedings of the National Academy of Sciences*, 109(26):10444–10449.

Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D., and Veitia, R. A. (2010). Heterosis. *The Plant Cell*, 22(7):2105–2112.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15):2114–2120.

Borges, F. and Martienssen, R. A. (2015). The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell Biology*, 16(12):727–741. Number: 12 Publisher: Nature Publishing Group.

Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics (Oxford, England)*, 19(7):889–890.

Caro, E., Stroud, H., Greenberg, M. V. C., Bernatavichute, Y. V., Feng, S., Groth, M., Vashisht, A. A., Wohlschlegel, J., and Jacobsen, S. E. (2012). The SET-domain protein SUVR5 mediates H3K9me2 deposition and silencing at stimulus response genes in a DNA methylation-independent manner. *PLoS genetics*, 8(10):e1002995.

Catoni, M. and Cortijo, S. (2018). Chapter Four - EpiRILs: Lessons From Arabidopsis. In Mirouze, M., Bucher, E., and Gallusci, P., editors, *Advances in Botanical Research*, volume 88 of *Plant Epigenetics Coming of Age for Breeding Applications*, pages 87–116. Academic Press.

Chan, S. W.-L., Zilberman, D., Xie, Z., Johansen, L. K., Carrington, J. C., and Jacobsen, S. E. (2004). RNA Silencing Genes Control de Novo DNA Methylation. *Science*, 303(5662):1336–1336. Publisher: American Association for the Advancement of Science.

Chandler, V. L. (2007). Paramutation: From Maize to Mice. *Cell*, 128(4):641–645.

Chen, L., Zhu, Y., Ren, X., Yao, D., Song, Y., Fan, S., Li, X., Zhang, Z., Yang, S., Zhang, J., and Zhang, J. (2022). Heterosis and Differential DNA Methylation in Soybean Hybrids and Their Parental Lines. *Plants*, 11(9):1136. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Chen, Z. J. (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant Science*, 15(2):57–71. Publisher: Elsevier.

Chodavarapu, R. K., Feng, S., Ding, B., Simon, S. A., Lopez, D., Jia, Y., Wang, G.-L., Meyers, B. C., Jacobsen, S. E., and Pellegrini, M. (2012). Transcriptome and methylome interactions in rice hybrids. *Proceedings of the National Academy of Sciences*, 109(30):12040–12045.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2):80–92.

Colomé-Tatché, M., Cortijo, S., Wardenaar, R., Morgado, L., Lahouze, B., Sarazin, A., Etcheverry, M., Martin, A., Feng, S., Duvernois-Berthet, E., Labadie, K., Wincker, P., Jacobsen, S. E., Jansen, R. C., Colot, V., and Johannes, F. (2012). Features of the Arabidopsis recombination landscape resulting from the combined loss of sequence variation and DNA methylation. *Proceedings of the National Academy of Sciences*, 109(40):16240–16245. Publisher: Proceedings of the National Academy of Sciences.

Cortijo, S., Wardenaar, R., Colomé-Tatché, M., Gilly, A., Etcheverry, M., Labadie, K., Caillieux, E., Hospital, F., Aury, J.-M., Wincker, P., Roudier, F., Jansen, R. C., Colot, V., and Johannes, F. (2014). Mapping the Epigenetic Basis of Complex Traits. *Science*, 343(6175):1145–1148. Publisher: American Association for the Advancement of Science.

Crow, J. F. (1948). Alternative Hypotheses of Hybrid Vigor. *Genetics*, 33(5):477–487.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2):giab008.

Dapp, M., Reinders, J., Bédiée, A., Balsera, C., Bucher, E., Theiler, G., Granier, C., and Paszkowski, J. (2015). Heterosis and inbreeding depression of epigenetic Arabidopsis hybrids. *Nature Plants*, 1(7).

Diez, C. M., Roessler, K., and Gaut, B. S. (2014). Epigenetics and plant genome evolution. *Current Opinion in Plant Biology*, 18:1–8.

East, E. M. (1936). Heterosis. *Genetics*, 21(4):375–397.

Fujimoto, R., Uezono, K., Ishikura, S., Osabe, K., Peacock, W. J., and Dennis, E. S. (2018). Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. *Breeding Science*, 68(2):145–158.

Furci, L., Jain, R., Stassen, J., Berkowitz, O., Whelan, J., Roquis, D., Baillet, V., Colot, V., Johannes, F., and Ton, J. (2019). Identification and characterisation of hypomethylated DNA loci controlling quantitative resistance in Arabidopsis. *eLife*, 8:e40655.

Ge, S. X., Jung, D., and Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 36(8):2628–2629.

Gendrel, A.-V., Lippman, Z., Yordan, C., Colot, V., and Martienssen, R. A. (2002). Dependence of Heterochromatic Histone H3 Methylation Patterns on the Arabidopsis Gene DDM1. *Science*, 297(5588):1871–1873. Publisher: American Association for the Advancement of Science.

Greaves, I., Groszmann, M., Dennis, E. S., and Peacock, W. J. (2012a). Trans-chromosomal methylation. *Epigenetics*, 7(8):800–805. Publisher: Taylor & Francis _eprint: https://doi.org/10.4161/epi.20820.

Greaves, I. K., Eichten, S. R., Groszmann, M., Wang, A., Ying, H., Peacock, W. J., and Dennis, E. S. (2016). Twenty-four–nucleotide siRNAs produce heritable trans-chromosomal methylation in F1 Arabidopsis hybrids. *Proceedings of the National Academy of Sciences of the United States of America*, 113(44):E6895–E6902.

Greaves, I. K., Groszmann, M., Wang, A., Peacock, W. J., and Dennis, E. S. (2014). Inheritance of Trans Chromosomal Methylation patterns from *Arabidopsis* F1 hybrids. *Proceedings of the National Academy of Sciences*, 111(5):2017–2022.

Greaves, I. K., Groszmann, M., Ying, H., Taylor, J. M., Peacock, W. J., and Dennis, E. S. (2012b). Trans Chromosomal Methylation in Arabidopsis hybrids. *Proceedings of the National Academy of Sciences*, 109(9):3570–3575. Publisher: Proceedings of the National Academy of Sciences.

Groszmann, M., Greaves, I. K., Albert, N., Fujimoto, R., Helliwell, C. A., Dennis, E. S., and Peacock, W. J. (2011a). Epigenetics in plants—vernalisation and hybrid vigour. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1809(8):427–437.

Groszmann, M., Greaves, I. K., Albertyn, Z. I., Scofield, G. N., Peacock, W. J., and Dennis, E. S. (2011b). Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. *Proceedings of the National Academy of Sciences*, 108(6):2617–2622.

Groszmann, M., Greaves, I. K., Fujimoto, R., James Peacock, W., and Dennis, E. S. (2013). The role of epigenetics in hybrid vigour. *Trends in Genetics*, 29(12):684–690.

Hazarika, R. R., Serra, M., Zhang, Z., Zhang, Y., Schmitz, R. J., and Johannes, F. (2022). Molecular properties of epimutation hotspots. *Nature Plants*, 8(2):146–156. Number: 2 Publisher: Nature Publishing Group.

He, G., Zhu, X., Elling, A. A., Chen, L., Wang, X., Guo, L., Liang, M., He, H., Zhang, H., Chen, F., Qi, Y., Chen, R., and Deng, X.-W. (2010). Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and Their Reciprocal Hybrids. *The Plant Cell*, 22(1):17–33.

He, L., Zhao, C., Zhang, Q., Zinta, G., Wang, D., Lozano-Durán, R., and Zhu, J.-K. (2021). Pathway conversion enables a double-lock mechanism to maintain DNA methylation and genome stability. *Proceedings of the National Academy of Sciences*, 118(35):e2107320118.

Hollister, J. D. and Gaut, B. S. (2009). Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research*, 19(8):1419–1428. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

Howard, M. M., Bae, A., Pirani, Z., Van, N., and Königer, M. (2020). Impairment of chloroplast movement reduces growth and delays reproduction of Arabidopsis thaliana in natural and controlled conditions. *American Journal of Botany*, 107(9):1309–1318. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ajb2.1537.

Hövel, I., Pearson, N. A., and Stam, M. (2015). Cis-acting determinants of paramutation. *Seminars in Cell & Developmental Biology*, 44:22–32.

Johannes, F., Porcher, E., Teixeira, F. K., Saliba-Colombani, V., Simon, M., Agier, N., Bulski, A., Albuisson, J., Heredia, F., Audigier, P., Bouchez, D., Dillmann, C., Guerche, P., Hospital, F., and Colot, V. (2009). Assessing the Impact of Transgenerational Epigenetic Variation on Complex Traits. *PLOS Genetics*, 5(6):e1000530. Publisher: Public Library of Science.

Johnson, L. M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., and Jacobsen, S. E. (2007). The SRA Methyl-Cytosine-Binding Domain Links DNA and Histone Methylation. *Current Biology*, 17(4):379–384.

Johnson, N. R., Yeoh, J. M., Coruh, C., and Axtell, M. J. (2016). Improved Placement of Multi-mapping Small RNAs. *G3 (Bethesda, Md.)*, 6(7):2103–2111.

Jones, D. F. (1917). Dominance of Linked Factors as a Means of Accounting for Heterosis. *Proceedings of the National Academy of Sciences*, 3(4):310–312. Publisher: Proceedings of the National Academy of Sciences.

Junker, A., Muraya, M. M., Weigelt-Fischer, K., Arana-Ceballos, F., Klukas, C., Melchinger, A. E., Meyer, R. C., Riewe, D., and Altmann, T. (2015). Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Frontiers in Plant Science*, 5.

Kakoulidou, I. and Johannes, F. (2023). DNA methylation remodeling in F1 hybrids. *The Plant Journal: For Cell and Molecular Biology*.

Kakoulidou, I., Piecyk, R. S., Meyer, R. C., Kuhlmann, M., Gutjahr, C., Altmann, T., and Johannes, F. (2024). Mapping parental DMRs predictive of local and distal methylome remodeling in epigenetic F1 hybrids. *Life Science Alliance*, 7(4):e202402599.

Kakutani, T., Jeddeloh, J. A., Flowers, S. K., Munakata, K., and Richards, E. J. (1996). Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proceedings of the National Academy of Sciences*, 93(22):12406–12411.

Kakutani, T., Jeddeloh, J. A., and Richards, E. J. (1995). Characterization of an Arabidopsis thaliana DNA hypomethylation mutant. *Nucleic Acids Research*, 23(1):130–137.

Kawanabe, T., Ishikura, S., Miyaji, N., Sasaki, T., Wu, L. M., Itabashi, E., Takada, S., Shimizu, M., Takasaki-Yasuda, T., Osabe, K., Peacock, W. J., Dennis, E. S., and Fujimoto, R. (2016). Role of DNA methylation in hybrid vigor in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, 113(43):E6704–E6711.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36.

Kinoshita, Y., Saze, H., Kinoshita, T., Miura, A., Soppe, W. J. J., Koornneef, M., and Kakutani, T. (2007). Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *The Plant Journal*, 49(1):38–45.

Klukas, C., Chen, D., and Pape, J.-M. (2014). Integrated Analysis Platform: An Open-Source Information System for High-Throughput Plant Phenotyping. *Plant Physiology*, 165(2):506–518.

Kooke, R., Johannes, F., Wardenaar, R., Becker, F., Etcheverry, M., Colot, V., Vreugdenhil, D., and Keurentjes, J. J. (2015). Epigenetic Basis of Morphological Variation and Phenotypic Plasticity in Arabidopsis thaliana. *The Plant Cell*, 27(2):337–348.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359.

Latzel, V., Allan, E., Bortolini Silveira, A., Colot, V., Fischer, M., and Bossdorf, O. (2013). Epigenetic diversity increases the productivity and stability of plant populations. *Nature Communications*, 4(1):2875. Number: 1 Publisher: Nature Publishing Group.

Lauss, K., Wardenaar, R., Oka, R., Hulten, M. H. A. v., Guryev, V., Keurentjes, J. J. B., Stam, M., and Johannes, F. (2018). Parental DNA Methylation States Are Associated with Heterosis in Epigenetic Hybrids. *Plant Physiology*, 176(2):1627–1645.

Law, J. A. and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, 11(3):204–220.

Li, H., Yuan, J., Wu, M., Han, Z., Li, L., Jiang, H., Jia, Y., Han, X., Liu, M., Sun, D., Chen, C., Song, W., and Wang, C. (2018). Transcriptome and DNA methylome reveal insights into yield heterosis in the curds of broccoli (Brassica oleracea L var. italic). *BMC Plant Biology*, 18(1):168.

Li, Y., Varala, K., Moose, S. P., and Hudson, M. E. (2012). The Inheritance Pattern of 24 nt siRNA Clusters in Arabidopsis Hybrids Is Influenced by Proximity to Transposable Elements. *PLOS ONE*, 7(10):e47043. Publisher: Public Library of Science.

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, 30(7):923–930.

Lippman, Z., Gendrel, A.-V., Black, M., Vaughn, M. W., Dedhia, N., Richard McCombie, W., Lavine, K., Mittal, V., May, B., Kasschau, K. D., Carrington, J. C., Doerge, R. W., Colot, V., and Martienssen, R. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature*, 430(6998):471–476. Number: 6998 Publisher: Nature Publishing Group.

Ma, X., Xing, F., Jia, Q., Zhang, Q., Hu, T., Wu, B., Shao, L., Zhao, Y., Zhang, Q., and Zhou, D.-X. (2021). Parental variation in CHG methylation is associated with allelic-specific expression in elite hybrid rice. *Plant Physiology*.

Marí-Ordóñez, A., Marchais, A., Etcheverry, M., Martin, A., Colot, V., and Voinnet, O. (2013). Reconstructing de novo silencing of an active plant retrotransposon. *Nature Genetics*, 45(9):1029–1039. Number: 9 Publisher: Nature Publishing Group.

Mathieu, O., Reinders, J., Čaikovski, M., Smathajitt, C., and Paszkowski, J. (2007). Transgenerational Stability of the Arabidopsis Epigenome Is Coordinated by CG Methylation. *Cell*, 130(5):851–862.

Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A. J. (2009). RNA-mediated chromatin-based silencing in plants. *Current Opinion in Cell Biology*, 21(3):367–376.

Matzke, M. A. and Mosher, R. A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics*, 15(6):394–408. Number: 6 Publisher: Nature Publishing Group.

Melchinger, A. E., Piepho, H.-P., Utz, H. F., Muminović, J., Wegenast, T., Törjék, O., Altmann, T., and Kusterer, B. (2007). Genetic Basis of Heterosis for Growth-Related Traits in Arabidopsis Investigated by Testcross Progenies of Near-Isogenic Lines Reveals a Significant Role of Epistasis. *Genetics*, 177(3):1827–1837.

Meng, D., Dubin, M., Zhang, P., Osborne, E. J., Stegle, O., Clark, R. M., and Nordborg, M. (2016). Limited Contribution of DNA Methylation Variation to Expression Regulation in Arabidopsis thaliana. *PLoS Genetics*, 12(7):e1006141.

Meyer, R. C., Törjék, O., Becher, M., and Altmann, T. (2004). Heterosis of Biomass Production in Arabidopsis. Establishment during Early Development. *Plant Physiology*, 134(4):1813–1823.

Mollá-Morales, A., Sarmiento-Mañús, R., Robles, P., Quesada, V., Pérez-Pérez, J. M., González-Bayón, R., Hannah, M. A., Willmitzer, L., Ponce, M. R., and Micol, J. L. (2011). Analysis of ven3 and ven6 reticulate mutants reveals the importance of arginine biosynthesis in Arabidopsis leaf development. *The Plant Journal: For Cell and Molecular Biology*, 65(3):335–345.

Ni, Z., Kim, E.-D., Ha, M., Lackey, E., Liu, J., Zhang, Y., Sun, Q., and Chen, Z. J. (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature*, 457(7227):327–331.

Oikawa, K., Kasahara, M., Kiyosue, T., Kagawa, T., Suetsugu, N., Takahashi, F., Kanegae, T., Niwa, Y., Kadota, A., and Wada, M. (2003). CHLOROPLAST UNUSUAL POSITIONING1 Is Essential for Proper Chloroplast Positioning. *The Plant Cell*, 15(12):2805–2815.

Quadrana, L., Etcheverry, M., Gilly, A., Caillieux, E., Madoui, M.-A., Guy, J., Bortolini Silveira, A., Engelen, S., Baillet, V., Wincker, P., Aury, J.-M., and Colot, V. (2019). Transposition favors the generation of large effect mutations that may facilitate rapid adaption. *Nature Communications*, 10(1):3421. Number: 1 Publisher: Nature Publishing Group.

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6):841–842.

Reinders, J., Wulff, B. B. H., Mirouze, M., Marí-Ordóñez, A., Dapp, M., Rozhon, W., Bucher, E., Theiler, G., and Paszkowski, J. (2009). Compromised stability of DNA methylation and transposon immobilization in mosaic Arabidopsis epigenomes. *Genes & Development*, 23(8):939–950.

Rigal, M., Becker, C., Pélissier, T., Pogorelcnik, R., Devos, J., Ikeda, Y., Weigel, D., and Mathieu, O. (2016). Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in Arabidopsis thaliana F1 epihybrids. *Proceedings of the National Academy of Sciences*, 113(14):E2083–E2092.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1):24–26.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, 26(1):139–140.

Roux, F., Colomé-Tatché, M., Edelist, C., Wardenaar, R., Guerche, P., Hospital, F., Colot, V., Jansen, R. C., and Johannes, F. (2011). Genome-Wide Epigenetic Perturbation Jump-Starts Patterns of Heritable Variation Found in Nature. *Genetics*, 188(4):1015–1017.

Saze, H., Scheid, O. M., and Paszkowski, J. (2003). Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nature Genetics*, 34(1):65–69. Number: 1 Publisher: Nature Publishing Group.

Schnable, P. S. and Springer, N. M. (2013). Progress Toward Understanding Heterosis in Crop Plants. *Annual Review of Plant Biology*, 64(1):71–88.

Secco, D., Wang, C., Shou, H., Schultz, M. D., Chiarenza, S., Nussaume, L., Ecker, J. R., Whelan, J., and Lister, R. (2015). Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *eLife*, 4:e09343. Publisher: eLife Sciences Publications, Ltd.

Seifert, F., Thiemann, A., Grant-Downton, R., Edelmann, S., Rybka, D., Schrag, T. A., Frisch, M., Dickinson, H. G., Melchinger, A. E., and Scholten, S. (2018a). Parental Expression Variation of Small RNAs Is Negatively Correlated with Grain Yield Heterosis in a Maize Breeding Population. *Frontiers in Plant Science*, 9.

Seifert, F., Thiemann, A., Schrag, T. A., Rybka, D., Melchinger, A. E., Frisch, M., and Scholten, S. (2018b). Small RNA-based prediction of hybrid performance in maize. *BMC Genomics*, 19(1):371.

Shahryary, Y., Hazarika, R. R., and Johannes, F. (2020a). MethylStar: A fast and robust pre-processing pipeline for bulk or single-cell whole-genome bisulfite sequencing data. *BMC Genomics*, 21(1):479.

Shahryary, Y., Symeonidi, A., Hazarika, R. R., Denkena, J., Mubeen, T., Hofmeister, B., van Gurp, T., Colomé-Tatché, M., Verhoeven, K. J., Tuskan, G., Schmitz, R. J., and Johannes, F. (2020b). AlphaBeta: computational inference of epimutation rates and spectra from high-throughput DNA methylation data in plants. *Genome Biology*, 21(1):260.

Shen, H., He, H., Li, J., Chen, W., Wang, X., Guo, L., Peng, Z., He, G., Zhong, S., Qi, Y., Terzaghi, W., and Deng, X. W. (2012). Genome-Wide Analysis of DNA Methylation and Gene Expression Changes in Two Arabidopsis Ecotypes and Their Reciprocal Hybrids. *The Plant Cell*, 24(3):875–892.

Shen, Y., Sun, S., Hua, S., Shen, E., Ye, C.-Y., Cai, D., Timko, M. P., Zhu, Q.-H., and Fan, L. (2017). Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid Brassica napus uncovers key roles for small RNAs. *The Plant Journal*, 91(5):874–893.

Shivaprasad, P. V., Dunn, R. M., Santos, B. A., Bassett, A., and Baulcombe, D. C. (2012). Extraordinary transgressive phenotypes of hybrid tomato are influenced by epigenetics and small silencing RNAs. *The EMBO Journal*, 31(2):257–266.

Sinha, P., Singh, V. K., Saxena, R. K., Kale, S. M., Li, Y., Garg, V., Meifang, T., Khan, A. W., Kim, K. D., Chitikineni, A., Saxena, K. B., Sameer Kumar, C. V., Liu, X., Xu, X., Jackson, S., Powell, W., Nevo, E., Searle, I. R., Lodha, M., and Varshney, R. K. (2020). Genome-wide analysis of epigenetic and transcriptional changes associated with heterosis in pigeonpea. *Plant Biotechnology Journal*, 18(8):1697–1710. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/pbi.13333.

Soppe, W. J. J., Jacobsen, S. E., Alonso-Blanco, C., Jackson, J. P., Kakutani, T., Koornneef, M., and Peeters, A. J. M. (2000). The Late Flowering Phenotype of fwa Mutants Is Caused by Gain-of-Function Epigenetic Alleles of a Homeodomain Gene. *Molecular Cell*, 6(4):791–802.

Team, R. C. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

To, T. K. and Kakutani, T. (2022). Crosstalk among pathways to generate DNA methylome. *Current Opinion in Plant Biology*, 68:102248.

Vongs, A., Kakutani, T., Martienssen, R. A., and Richards, E. J. (1993). Arabidopsis thaliana DNA Methylation Mutants. *Science*, 260(5116):1926–1928. Publisher: American Association for the Advancement of Science.

Williams, B. P. and Gehring, M. (2020). Principles of Epigenetic Homeostasis Shared Between Flowering Plants and Mammals. *Trends in Genetics*, 36(10):751–763.

Yang, D.-L., Zhang, G., Tang, K., Li, J., Yang, L., Huang, H., Zhang, H., and Zhu, J.-K. (2016). Dicer-independent RNA-directed DNA methylation in Arabidopsis. *Cell Research*, 26(1):66–82. Number: 1 Publisher: Nature Publishing Group.

Yi, H. and Richards, E. J. (2008). Phenotypic instability of Arabidopsis alleles affecting a disease Resistancegene cluster. *BMC Plant Biology*, 8(1):36.

Zemach, A., Kim, M. Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S. L., and Zilberman, D. (2013). The Arabidopsis Nucleosome Remodeler DDM1 Allows DNA Methyltransferases to Access H1-Containing Heterochromatin. *Cell*, 153(1):193–205.

Zhang, C., Pang, Q., Jiang, L., Wang, S., Yan, X., Chen, S., and He, Y. (2015). Dihydroxy-acid dehydratase is important for gametophyte development and disruption causes increased susceptibility to salinity stress in Arabidopsis. *Journal of Experimental Botany*, 66(3):879–888.

Zhang, Q., Wang, D., Lang, Z., He, L., Yang, L., Zeng, L., Li, Y., Zhao, C., Huang, H., Zhang, H., Zhang, H., and Zhu, J.-K. (2016). Methylation interactions in Arabidopsis hybrids require RNA-directed DNA methylation and are influenced by genetic variation. *Proceedings of the National Academy of Sciences*, 113(29):E4248–E4256.

Zhang, Y., Jang, H., Xiao, R., Kakoulidou, I., Piecyk, R. S., Johannes, F., and Schmitz, R. J. (2021). Heterochromatin is a quantitative trait associated with spontaneous epiallele formation. *Nature Communications*, 12(1):6958. Number: 1 Publisher: Nature Publishing Group.

Zhou, M., Palanca, A. M. S., and Law, J. A. (2018). Locus-specific control of the de novo DNA methylation pathway in Arabidopsis by the CLASSY family. *Nature Genetics*, 50(6):865–873. Number: 6 Publisher: Nature Publishing Group.

Acknowledgments

Completing this Ph.D. has been a journey of learning and growth, and there are many who have contributed to this process. Foremost, I would like to express my gratitude to Pavlos, for giving me the first insights into the world of academia, and to Frank, my supervisor, for providing me with the opportunity to pursue this challenge.

I am grateful to the entire JLab group, whose company has made the last years more enjoyble. A special mention to Katia, whose availability and support have been invaluable. To Rashmi and Yadi, my initial companions in the group; To Robert, whose help was crucial in the success of this project; To Lucas, for his ability to bring positivity when it was most needed, and to Zhillin, for the laughter and smiles that brightened our days. Also, I wanna thank Claus and Caroline for their welcoming support in their groups during my time with them and above all, a special acknowledgment is reserved for Susanna, 'the best secretary ever'. Her presence have been a source of comfort and motivation, truly making her the person everyone hopes to have by their side.

Furthermore, I would like to thank the entire team at IPK for a year that will forever be etched in my memory. To Thomas, for initiating this collaboration; To Rhonda, for guiding me during my first steps in this world; To Monika, an increadible woman that took care of me from day one and to the entire group that made that year an extraordinary chapter of my life.

Finally, I can't thank enough my parents for shaping me into the person I am today. Their love and guidance have been my foundation. And above all, to Apostolos, who has been with me on this journey since the beginning, and to whom I owe the decision of undertaking this challenge at first place.