

Analysis of the effect of DNA methylation on
heterosis formation in *Zea mays* and *Brassica napus*

Dissertation

with the aim of achieving a doctoral degree
at the Faculty of Mathematics, Informatics and Natural Sciences
Department of Biology
of the University of Hamburg

submitted by
Susanne Edelmann

Hamburg, May 2019

Supervisor: Prof. Dr. Stefan Scholten

1st Examiner: Prof. Dr. Stefan Scholten

2nd Examiner: Prof. Dr. Arp Schnittger

Head of examination committee: Prof. Dr. Stefan Hoth

Disputation: 16/08/2019

This work has been generated from September 2014 until May 2019 in the lab of Prof. Dr. Stefan Scholten in the group Developmental Biology in the department of Biology at the University of Hamburg, Institute of Plant Science and Micobiology, and in the group Quantitative Genetics and Genomics of Crops in the institute of Plant Breeding, Seed Science and Population Genetics at the University of Hohenheim.

Oral and poster presentations, covering parts of this work:

57th Maize Genetics Conference 2015, 12.03. - 15.03.2015, St. Charles, USA, Poster presentation

Epigenetic & Chromatin Regulation of Plant Traits, 14.01. - 15.01.2016, Strasbourg, France, Poster presentation

European Molecular Maize Meeting 2016, 18.05. - 20.05.2016, Hamburg, Germany, Poster presentation

33th Interdisciplinary Plant Group Symposium, 25.05. - 27.05.2016, Columbia, USA, Oral presentation

European Molecular Maize Meeting 2017, 03.05. - 05.05.2017, Ghent, Belgium, Poster presentation

5th European Workshop on Plant Chromatin, 01.06. - 02.06.2017, Vienna, Austria, Short talk and Poster presentation

60th Maize Genetics Conference 2018, 22.03. - 25.03.2018, Saint-Malo, France, Poster presentation

Publication prepared during graduation and at least partly covered in this work:

Edelmann S, Scholten S. 2018. Bisulfite Sequencing Using Small DNA Amounts. **In:** **Bemer M, Baroux C, eds.** Plant Chromatin Dynamics. *Springer: Methods in Molecular Biology* **1675**: 45-61

Abstract

Heterosis or hybrid vigor is the superior performance of F1 hybrid plants over their parental inbred lines. Heterosis is particularly important in the production systems of major crops. Recent studies have suggested an epigenetic contribution such as DNA methylation to heterosis, but the molecular mechanism underlying heterosis still remains unclear. To assess whether developmental timing of epigenetic regulation plays a role in heterosis in maize, early embryos were dissected and treated with the demethylating agent 5-aza-2'-deoxycytidine (aza) in an *in vitro* culture system. Reduced representation bisulfite sequencing confirmed successful hypomethylation of DNA induced in the early embryo and maintained to the seedling stage. Comparing seedling growth rate heterosis between plants derived from aza-treated embryos and their non-treated control group demonstrated a significant increase of growth heterosis upon demethylation, indicating that DNA methylation patterns established during early embryogenesis of maize play a role in seedling growth heterosis formation.

The impact of DNA methylation on heterosis formation and hybrid performance in a broad range of traits was determined in a subset of a *Brassica napus* population. DNA methylation patterns between maternal, paternal and hybrid lines revealed a trend towards hypomethylation of repetitive regions in the hybrid. In parental differentially methylated regions, methylation levels of hybrids revealed a majority of dominant methylation patterns, where the hybrid methylation level adapted to either one of the parental alleles. Trans-chromosomal demethylation events were observed especially for CHH contexts. Coincidentally with a higher methylation level, a decrease in mRNA expression was revealed in the hybrid for genes active in RdDM in the hybrid. Certain stress-induced genes were down-regulated together with a higher methylation level, especially in the salicylic acid pathway. Together, the results indicate that hypomethylation of transposable elements in *Brassica napus* hybrids potentially lead to higher expression of adjacent genes. Moreover, the basal plant stress response is down-regulated in hybrids, which might impair growth limiting factors. Both these mechanisms might promote heterosis.

Zusammenfassung

Heterosis ist die verbesserte Leistung von heterozygoten Nachkommen im Vergleich zu den parental homozygoten Inzuchtlinien und wird unter anderem mit epigenetischen Modifikationen wie DNA-Methylierung in Verbindung gebracht. Erstes Ziel dieser Arbeit war es, eine Verbindung zwischen Heterosis und DNA-Methylierung während der frühen Embryogenese von Mais herzustellen. Hierfür wurden Embryonen von Inzuchtlinien und Hybriden von Mais in der frühen Embryonalentwicklung mit dem Methyltransferaseinhibitor 5-Aza-2'-deoxycytidin (Aza) behandelt. Durch die Aza-Behandlung erfolgte eine Demethylierung der Embryo-DNA, die über Bisulfitsequenzierung von randomisierten repräsentativen Genombereichen nachgewiesen werden konnte und während der Entwicklung zum Keimling erhalten blieb. Durch die Demethylierung wurde eine signifikante Erhöhung der Wachstumsheterosis von aus Aza-behandelten Embryonen generierten Pflanzen im Vergleich zu der Kontrollgruppe hervorgerufen. Eine mögliche Schlussfolgerung aus diesen Ergebnissen ist, dass die DNA-Methylierung in der frühen Embryonalentwicklung eine kontrollierende bzw. zügelnde Wirkung auf Heterosis ausübt.

Der Einfluss von DNA-Methylierung auf Heterosis wurde zudem im größeren Umfang in einer *Brassica napus* Population untersucht. Bei der Charakterisierung der Methylierungsmuster zwischen maternalen, paternalen und Hybridlinien konnte ein Trend in Richtung Hypomethylierung der repetitiven Genombereiche im Hybriden festgestellt werden. In zwischen den Eltern differentiell methylierten Regionen passten sich die Methylierungslevel der Hybride überwiegend an das Methylierungsniveau eines der Eltern an. Ein größerer Anteil der CHH-Methylierung näherte sich dem Allel des niedriger methylierten Elters an. Die Expression einiger Gene, die von *de novo* Methylierungsmechanismen bekannt sind, war vermindert. Gene, die von der Salicylsäure-induzierten Stressantwort in Pflanzen bekannt sind, wurden im Hybrid herunterreguliert, was gemeinsam mit einem höheren Methylierungsniveau beobachtet wurde. Zusammen deuten die Ergebnisse an, dass Transposons in *Brassica napus* Hybriden weniger methyliert sind, was einen aktivierenden Effekt auf benachbarte Gene haben könnte. Zudem ist auch die pflanzliche Stressantwort im Ruhezustand im Hybriden heruntergefahren, wodurch limitierende Faktoren inhibiert sein könnten. Diese beiden Mechanismen könnten einen Beitrag zu Heterosis leisten.

Table of Content

Abstract	VII
Zusammenfassung	IX
Index of Figures	XV
Index of Tables	XIX
Abbreviation Index	XXI
Unit Index	XXV
1 Introduction	1
1.1 Model plants	1
1.1.1 Maize.....	1
1.1.2 Rapeseed.....	2
1.2 Heterosis	2
1.3 Epigenetic mechanisms	8
1.4 Effects of DNA methylation on heterosis	10
1.5 Reduced Representation Bisulfite Sequencing.....	12
1.6 Objectives	13
2 Material and Methods	15
2.1 Materials	15
2.1.1 Plant material.....	15
2.1.2 Chemicals, solutions and media	16
2.1.3 Used oligonucleotides.....	18
2.1.4 Software.....	20
2.1.5 Data.....	21

2.2	Methods.....	22
2.2.1	Plant growth and cultivation	22
2.2.2	Reduced Representation Bisulfite Sequencing (RRBS).....	24
2.2.3	Computational methods.....	29
3	Results.....	33
3.1	Optimization of the RRBS protocol.....	33
3.1.1	gDNA digestion and size selection.....	33
3.1.2	Comparison of two RRBS protocols.....	36
3.2	DNA methylation impact on heterosis during early embryogenesis in maize	38
3.2.1	Confirmation of successful DNA demethylation through aza-treatment	39
3.2.2	Effects of the DNA demethylation during embryo development on seedling growth rate	41
3.3	DNA methylation patterns in an oilseed rape population	43
3.3.1	Phenotypic variation in the <i>Brassica napus</i> population.....	43
3.3.2	Summary statistics of DNA methylation.....	49
3.3.3	Analysis of parental differential methylation.....	52
3.3.4	DNA methylation pattern analysis on genome annotations.....	57
3.3.5	DNA methylation and mRNA expression.....	62
3.3.6	DNA methylation and sRNA expression	64
3.3.7	Analysis of genes involved in regulatory pathways.....	67
3.3.8	Summary of DNA methylation and RNA expression characteristics.....	70
3.3.9	Heterosis-associated DMRs	75
4	Discussion.....	79
4.1	Reduced representation bisulfite sequencing method.....	79
4.2	Impact of DNA methylation on heterosis in maize and oilseed rape.....	81
4.2.1	Early embryo maize tissue culture	81
4.2.2	Population DNA methylation dynamics of <i>Brassica napus</i>	85
4.3	Factors contributing to Heterosis.....	99

5	Perspective.....	107
	References	109
A	Appendix	XXXI
A.1	Tables	XXXI
	Declaration on oath / Eidesstattliche Erklärung	XLIII
	Confirmation of correct English	XLV
	Acknowledgement	XLVII

Index of Figures

Figure 1.1: Plants of the inbred lines UH301 (301x301, left) and UH005 (005x005, right) and their reciprocal hybrids (301x005, 005x301, the two central plants). The hybrids exhibit heterosis for plant height exceeding their parents (Meyer, 2007).....	1
Figure 1.2: The U-triangle shows chromosome relations between the <i>Brassica</i> species (Nagaharu (1935), modified by Purty <i>et al.</i> (2008)).	2
Figure 1.3: Model of trade-off between hybrid vigor and stress response gene expression proposed by Fujimoto <i>et al.</i> (2018).	6
Figure 1.4: Model of hybrid vigor proposed by Bar-Zvi <i>et al.</i> (2017).	7
Figure 1.5: Methylation pathways of the symmetric (CG and CNG, where N is similar to H) and asymmetric (CHH) contexts in <i>Arabidopsis thaliana</i> (Chan <i>et al.</i> , 2005).....	8
Figure 1.6: The maintenance methylation and RdDM pathway in plants (Fujimoto <i>et al.</i> , 2018).....	9
Figure 1.7: Principle of conversion of unmethylated cytosines to uracil through bisulfite (HSO ₃ ⁻) treatment. Unmethylated cytosines are deaminated to uracil, methylated cytosines are not converted (Krueger <i>et al.</i> , 2012).	12
Figure 2.1: Process of the nucellus in vitro culture.	23
Figure 2.2: Seedlings of <i>Brassica napus</i> genotypes at time of sampling.	24
Figure 2.3: Schematic overview over the RRBS library preparation method.	25
Figure 2.4: Structure of a custom column to elute DNA from an agarose gel.	27
Figure 3.1: Digestion with different enzymes and their respective combinations.	34
Figure 3.2: Gelectrophoretic separation of Taq ^α I restricted <i>Brassica napus</i> gDNA.	35
Figure 3.3: <i>In silico</i> restriction and fragment size selection of the used species.	36
Figure 3.4: FastQC (v0.11.4) analysis of RRBS sequence reads (performed by Dr. Felix Seifert).....	38

Figure 3.5: Methylation distribution patterns of aza-treated and non-treated embryonic DNA.	40
Figure 3.6: Growth rates and heterosis calculations of maize inbred lines and hybrids grown from aza-treated and non-treated embryos (1 dap).	42
Figure 3.7: Phenotypic variability of field traits.	45
Figure 3.8: Phenotypic variability of <i>in vitro</i> seedling root lengths.....	47
Figure 3.9: Phenotypic variability of <i>in vitro</i> seedling root lengths.....	48
Figure 3.10: Methylation rates of paternal lines, maternal lines and hybrids.....	50
Figure 3.11: Correlation of cytosine coverage and mean DNA methylation rates.	51
Figure 3.12: Number of differentially methylated positions per difference threshold for each cytosine context and hybrid line.	52
Figure 3.13: Number of differentially methylated regions per difference threshold for each cytosine context and hybrid line.....	53
Figure 3.14: Heatmap based on Pearson coefficient shows correlations between means and numbers (up and down regulation and both) of the differentially methylated positions and regions of the three contexts.....	54
Figure 3.15: DNA methylation patterns in hybrids compared to parents in regions of parental differential methylation of all three cytosine contexts.	55
Figure 3.16: Adaption of the hybrid methylation level to either the maternal or paternal state in regions of parental differential methylation in hybrids of all three cytosine contexts.	56
Figure 3.17: Trans-chromosomal DNA methylation (TCM) and demethylation (TCdM) events in regions of parental differential methylation in hybrids of all three cytosine contexts.	56
Figure 3.18: Cytosine distribution on CDS, TEs and remaining regions.	57
Figure 3.19: Cytosine methylation rate distribution in 100 bins over CDS including 1kb up- and downstream regions of <i>Brassica napus</i>	58
Figure 3.20: Cytosine methylation rate distribution in 100 bins over TEs including 1kb up- and downstream regions of <i>Brassica napus</i>	59

Figure 3.21: Heatmap based on Pearson correlation coefficients of field traits of mature plants and root and shoot seedling traits.	61
Figure 3.22: Heatmap based on Pearson coefficient shows correlations between DMR methylation means and DMR numbers (total, on CDS, on TEs) of the three contexts with field traits of mature plants and root and shoot seedling traits.	62
Figure 3.23: Correlation of DNA methylation rates per contexts with mRNA expression.....	63
Figure 3.24: Expression patterns of transcripts of hybrids compared to parental lines relative to the total number of hybrid transcripts overlapping with parental DEGs.	63
Figure 3.25: Overlap of parental DEGs with parental DMRs.	64
Figure 3.26: sRNA distribution on features that are exposed with cytosines covered in methylation data.	65
Figure 3.27: Characterization of mapped sRNAs by size and expression.	65
Figure 3.28: Correlation of sRNA expression with DNA methylation rates (in %) of CG, CHG and CHH contexts.	66
Figure 3.29: Heatmap based on Pearson correlation coefficients between sRNA expression on DNA methylation and DMR loci (total, CDS, TEs) of the three cytosine contexts with field traits of mature plants and root and shoot seedling traits of hybrids.	67
Figure 3.30: Heatmap based on Pearson correlation coefficients shown for DNA methylation rate, sRNA and mRNA expression means of selected gene loci with field traits of mature plants and root/shoot seedling traits.	69
Figure 3.31: mRNA expression in DEGs and CHH methylation at sRNA locations in TEs shown as representative traits significantly differing between parents and hybrids.	70
Figure 3.32: sRNA expression levels shown as representative traits significantly differ between low performing and high performing hybrids.	72
Figure 3.33: Characterization of heterosis-associated CG DMRs and correlated features.	77

Figure 3.34: Characterization of heterosis-associated CHG DMRs and correlated features.	78
Figure 4.1: Hypothetical model illustrating some of the contributors to heterosis formation.	104

Index of Tables

Table 2.1: Breeding scheme of the rapeseed population.....	16
Table 2.2: Prepared solutions for <i>in vitro</i> culture.....	17
Table 2.3: Solutions for gDNA extraction.	17
Table 2.4: Prepared media for <i>in vitro</i> culture.	18
Table 2.5: Oligonucleotide sequences.....	18
Table 2.6: Software used in this study.....	20
Table 3.1: Summary of the sequencing results before and after adapter trimming.	37
Table 3.2: Summary of the sequencing read quality and quantity of maize samples. ..	39
Table 3.3: Averages and standard deviations of the analyzed phenotypic <i>in vitro</i> seedling root/shoot as well as field traits per maternal lines, paternal lines and hybrids.....	44
Table 3.4: Summary of the sequencing read quantity as well as methylation statistics for sequenced samples of rapeseed maternal lines, paternal lines and hybrids.....	49
Table 3.5: Summary of mRNA and sRNA expression as well as DNA methylation changes of parents vs. hybrids.....	68
Table 3.6: Summary of mRNA and sRNA expression as well as DNA methylation changes of parents vs. hybrids.....	71
Table 3.7: Summary of mRNA and sRNA expression as well as DNA methylation changes divided by cytosine context of low performing hybrids vs. high performing hybrids in the phenotypic trait MPH of seedling total root growth rate.....	73
Table 3.8: Summary of hybrid DNA methylation changes in regions of parental DMRs of low performing hybrids vs. high performing hybrids in the phenotypic trait MPH of seedling total root growth rate.....	75
Table A.1: Phenotypic field data of the Core factorial.	XXXI

Table A.2: Phenotypic <i>in vitro</i> seedling root and shoot data of the Core factorial. .	XXXIII
Table A.3: Gene loci of <i>Brassica napus</i> genes used for comparison with DNA methylation, mRNA and sRNA data.	XXXV
Table A.4: Analysis of heterosis for seedling growth rate for hybrids (A188 x H99) and inbred lines (A188 and H99) grown <i>in vitro</i> from aza-treated and non-treated embryos.....	XXXV
Table A.5: Basic statistics of RRBS data of the <i>Brassica napus</i> "Core factorial". Methylation (CpG, CHG, CHH, in %) and coverage (Ccov CG, Ccov CHG, Ccov CHH, in %) of the three cytosine contexts were calculated with 5 reads/cytosine.....	XXXVI
Table A.6: DNA methylation levels (in %) of enzyme locations including 1000 bp up- and downstream of each locus. Cytosine contexts are combined.....	XXXIX
Table A.7: mRNA expression levels (in FPKM) of enzyme locations including 1000 bp up- and downstream of each locus. Black indexes indicate hybrid lines, red maternal and blue paternal lines.	XL
Table A.8: Location of heterosis-associated DMRs per context.....	XLI
Table A.9: Mean methylation rates given in literature of <i>Brassica napus</i> and its ancestors <i>Brassica rapa</i> and <i>Brassica oleracea</i>	XLII

Abbreviation Index

A	A	Adenine
	aza	5-aza-2'-deoxycytidine
B	BAP	Benzylaminopurin
	BPH	best-parent heterosis
C	C	Cytosine
	ca.	circa
	cDNA	copy DNA
	CDS	coding sequence
	CG	Cytosine guanine context
	CHG	Cytosine adenine/thymine/cytosine guanine context
	CHH	Cytosine adenine/thymine/cytosine adenine/thymine/cytosine context
	cv.	cultivar
D	dag	days after germination
	dap	days after pollination
	das	days after sowing
	ddH ₂ O	double-distilled water
	DNA	deoxyribonucleic acid
	DMR	differentially methylated region
	dsRNA	double-stranded RNA
	dNTP	deoxynucleotide triphosphate
E	e.g.	for example
	<i>E. coli</i>	<i>Escherichia coli</i>
	EDTA	Ethylenediaminetetraacetic acid
	<i>et al.</i>	<i>et alii</i>
	EtBr	Ethidium bromide
	etc.	<i>et cetera</i>

	EtOH abs.	96-100 % Ethanol
	EUE	Energy use efficiency
F	F ₁	first filial generation of offspring
	FDR	false discovery rate
	Fe-EDTA	Ferric ethylenediaminetetraacetic acid
G	gDNA	genomic deoxyribonucleic acid
	G	Guanine
H	ha-DMR	heterosis-associated differentially methylated region
	HPLC	high performance liquid chromatography
L	L.	Linné
M	max	maximal
	MPH	mid-parent heterosis
	miRNA	micro RNA
	mRNA	messenger RNA
	MS	Murashige-Skoog
	MSAP	Methylation sensitive amplification polymorphism
P	p	p-value
	PCR	Polymerase chain reaction
	Pol	DNA-dependent RNA-polymerase
R	r	correlation coefficient
	RdDM	RNA dependent DNA methylation
	RNA	ribonucleic acid
	RNAi	RNA interference
	RNase	ribonuclease
	rpm	rounds per minute
	RRBS	Reduced Representation Bisulfite Sequencing
S	SNP	single nucleotide polymorphism

sRNA	small RNA
siRNA	short interfering RNA
ssRNA	single stranded RNA
T	
T	Thymine
TAE	Tris-acetate-EDTA
ta-siRNA	trans-acting siRNA
TE	transposable element
TSS	transcription start site
TE buffer	Tris-EDTA buffer
Thiamine HCl	Thiamine hydrochloride
Tris-HCl	Tris(hydroxymethyl) aminomethane hydrochloride

Gene names are written in capital letters.

Unit Index

bp	base-pair
°C	Grad Celsius
d	days
h	hours
ha	hectar
L	liter
mol	molar mass
min	minutes
nt	nucleotide
FPKM	fragments per kilobase million
TPM	transcripts per million
rpmqn	read counts per million quantile normalized reads
kLux	kilo lux

This work is using SI units, all other units are listed in the unit index

1 Introduction

1.1 Model plants

1.1.1 Maize

Maize or corn (*Zea mays* L.) is one of the most important crops worldwide (Shiferaw *et al.*, 2011). *Zea mays subsp. mays* is mainly grown for production of animal feed, but is also used for biogas and food production (Ranum *et al.*, 2014). In research, maize is used as a model plant, as techniques like embryo or gamete isolation and controlled pollination are relatively uncomplicated compared to other cereals or smaller plants like *Arabidopsis thaliana* (Dumas, 1993). The maize genome has undergone large-scale duplication in its evolution (Adams & Wendel, 2005) and comprises 2.5 Gbp, including a large number of repetitive elements as a results of multiple transposon blooms (Schnable, 2015). Maize is usually grown in hybrid crosses, because of the heterosis effect. This phenomenon is well known in maize (Shull, 1908), because heterosis is especially pronounced (**Figure 1.1**). In general, heterosis describes the superior performance of hybrids compared to their parental inbred lines (East, 1936).

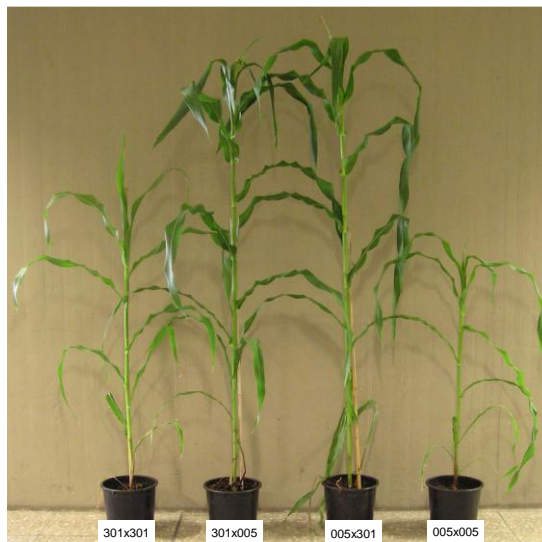


Figure 1.1: Plants of the inbred lines UH301 (301x301, left) and UH005 (005x005, right) and their reciprocal hybrids (301x005, 005x301, the two central plants). The hybrids exhibit heterosis for plant height exceeding their parents (Meyer, 2007).

1.1.2 Rapeseed

Rapeseed (*Brassica napus* L.) is one of the most important oilseed crops (Carré & Pouzet, 2014). It is used for vegetable oil production for human consumption, but also as animal feed and biofuel feedstocks. *Brassica napus* is an allotetraploid organism and has evolved ~7500 years ago by hybridization between *Brassica rapa* (A genome) and *Brassica oleracea* (C genome, **Figure 1.2**), followed by chromosome doubling (Chalhoub et al., 2014).

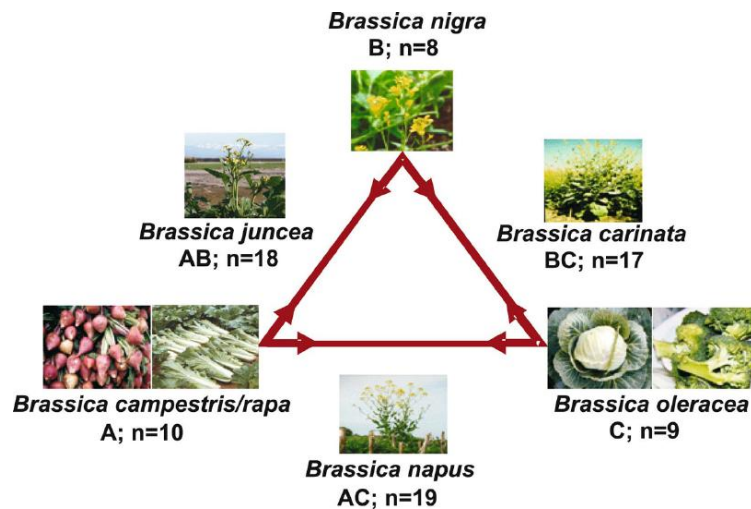


Figure 1.2: The U-triangle shows chromosome relations between the *Brassica* species (Nagaharu (1935), modified by Purty *et al.* (2008)).

The constituent A_n and C_n subgenomes are engaged in subtle structural, functional, and epigenetic cross-talk, with abundant homoeologous exchanges (Chalhoub *et al.*, 2014). Studies of polyploids soon after their formation have revealed genetic and epigenetic interactions between homologous genes. These interactions can be related to the phenotypes and evolutionary fates of polyploids (Comai, 2005). Most of the rapeseed grown are hybrid cultivars, which makes it like maize an important crop and suitable model plant to characterize hybrid performance and investigate the heterosis effect.

1.2 Heterosis

Heterosis, also termed hybrid vigor, is ubiquitous in nature, and is observed in all eukaryotic kingdoms including plants, animals and fungi (Herbst *et al.*, 2017). It describes

the superior phenotypic performance of heterozygous hybrids compared to their homozygous parental inbred lines (**Figure 1.1**). Heterosis has already been observed by Darwin in the 19th century in plants (Darwin, 1876). From the beginning of the 20th century cross pollination was used in maize to generate plants with enhanced performance (Shull, 1908). Heterosis has been exploited in plant and animal breeding in order to increase yield, size, biomass and other agriculturally important traits (Flint-Garcia *et al.*, 2007; Feng *et al.*, 2015).

As there are no new genes generated in the F1 hybrid, heterosis is likely caused by differences resulting from modifications of allelic or gene expression (Shen *et al.*, 2017). Several hypotheses have been proposed to explain the heterosis phenomenon, but they are not able to fully explain the genetic or molecular basis of heterosis (Schnable & Springer, 2013). The dominance model describes the complementation of different phenotypically deleterious alleles present in the inbred parental lines by superior alleles from the other parent in the hybrid (Crow, 1948). As it was not possible to breed inbred lines with exclusively advantageous alleles, this theory cannot be the only explanation (Swanson-Wagner *et al.* 2006). The second historical explanation for heterosis is the overdominance model, which refers to the idea that allelic interactions of key loci occur in the hybrid such that the heterozygous offspring performs better than either homozygous parent. This theory was confirmed for several loci (Schnable & Springer, 2013). An example is the SINGLE FLOWER TRUSS (SFT), which was shown to cause fruit yield hybrid vigor in tomato (*Solanum lycopersicum*), providing the first molecular evidence of heterosis formation caused by a single overdominantly expressed gene (Krieger *et al.*, 2010). Other explanations involved positive interactions of two alleles of different genes in *trans* called epistasis (Powers, 1944) or the repulsion-phase linkage of favorable and detrimental alleles mimicking overdominance (pseudo-overdominance, Moll *et al.*, 1964).

Heterosis has been widely analyzed on the molecular level over the last decades. Based on the modes of gene action in the hybrid, gene expression patterns have mainly been classified as additive, dominant and overdominant (non-additive). Additive expression represents mid-parental expression patterns in the hybrid, whereas dominance indicates either low or high parent-like expression. Overdominant gene expression levels in the hybrid are higher or lower than the parental level (Fujimoto *et al.*, 2018). Several studies analyzed the relation of transcription levels between hybrids and parents in plants (Guo *et al.*, 2006; He *et al.*, 2010; Meyer *et al.*, 2012). Some stud-

ies pointed towards the importance of the effect of non-additive gene expression between parent and hybrid on heterosis (Jahnke *et al.*, 2010; Fujimoto *et al.*, 2012; Zhu *et al.*, 2016), while others mainly detected additive gene expression in hybrids (Guo *et al.*, 2006; Swanson-Wagner *et al.*, 2006; Meyer *et al.*, 2012). In addition, both the additive and non-additive expression model were also observed at the same time (Stupar *et al.*, 2008). Although the modes of gene expression vary, the global trend of predominantly additive gene expression in hybrids compared to parental lines was commonly observed (reviewed in Feng *et al.*, 2015; Fujimoto *et al.*, 2018; Itabashi *et al.*, 2018). A heterotic phenotype can already be observed early in plant development (Jahnke *et al.*, 2010; Fujimoto *et al.*, 2012). Gene expression patterns, especially non-additive, vary in different developmental stages and tissues (Fujimoto *et al.*, 2012; Meyer *et al.*, 2012).

Genetic distance of parents was also correlated to heterosis; for example in *Brassica napus*, where three pools were generated from parental lines via marker polymorphisms. Crosses between the clusters led to higher yield heterosis than crosses from the same cluster (Riaz *et al.*, 2001). However, this correlation was not always found (Kawamura *et al.*, 2016; Yang *et al.*, 2017). Additionally, the genome dosage can influence heterosis, e.g. in triploid maize (Yao *et al.*, 2013).

Besides genetic factors, epigenetic factors and small RNAs (sRNAs) are also suggested to play a potential role in hybrid performance. Small RNAs, which belong to the group of non-protein-coding RNAs, are involved in transcriptional, post-transcriptional (e.g. RNA interference), and translational gene regulation by various mechanisms (reviewed in Zamore & Haley, 2005; Moazed, 2009; Castel & Martienssen, 2013). This includes involvement in epigenetic regulation (Aufsatz *et al.*, 2002; Zilberman *et al.*, 2002), thus they are likely playing a role in heterosis. In plants, sRNAs are classified in microRNA (miRNA) derived from single-stranded precursors with a hairpin structure and short-interfering RNA (siRNA) derived from double-stranded RNA precursors (Axtell, 2013). Several studies indicated a connection between sRNAs and heterosis: In maize seed, a number of miRNAs, that show significant expression in hybrids and parents, were expressed non-additively in the hybrid indicating a participation in heterosis (Ding *et al.*, 2012). Similar results were detected in rice hybrids (Zhang *et al.*, 2014). A predominant down-regulation of miRNAs and highly expressed sRNAs was observed in rice hybrids compared to inbred lines. Amongst them, some are predicted to target members of various developmental pathways indicating a contribution to heterosis (Chen *et al.*,

2010). In several studies, 24 nt siRNAs were shown to be down-regulated in hybrids compared to their parents assuming a role in heterosis (described in 1.4). Differential siRNA expression patterns linked to the RNA dependent DNA methylation pathway (RdDM) and differential methylation levels have been associated with heterosis (described in 1.4). Furthermore, histone modifications are known as epigenetic gene expression regulators (Dong & Weng, 2013). Although the histone modification patterns from the parents are largely passed to the F1 hybrid (Moghaddam *et al.*, 2011; Zhu *et al.*, 2017), a few loci show altered patterns correlating with gene expression (Ni *et al.*, 2009). Additionally, since histone modifications are linked to DNA methylation and sRNAs (Stroud *et al.*, 2014), a contribution to heterosis is likely.

Proteomic and metabolomic studies in hybrids found correlations between heterosis and protein abundance patterns, for instance in non-additive protein level variation, which is more frequent than non-additive gene expression variation in hybrids (Wang *et al.*, 2008). The same study showed that differential protein levels detected in heterotic hybrids are involved in stress response, protein and carbon metabolism (Wang *et al.*, 2008).

To combine the known impacts of genetic and epigenetic factors, dosage effects and gene expression on heterosis formation, Feng *et al.* (2015) proposed a possible model: Differential gene expression in the F1 hybrid induced by mixing the genomes of two distinct parents is mainly caused by epigenetic and genetic factors, and affected by genome dosage effects (e.g. in polyploids). Regulatory pathways, like the circadian clock (Shen *et al.*, 2012) and/or the energy-use efficiency (Hauben *et al.*, 2009; Goff, 2011), may be affected by these expression changes. This could lead to changes in downstream metabolic pathways, ultimately affecting various aspects of growth and development (Feng *et al.*, 2015).

Another theory was proposed by (Miller *et al.*, 2015) centering the balancing of differences between parental stress responses, which could be involved in biomass heterosis formation (**Figure 1.3**). The model is based upon hybrid necrosis, which can cause slow growth, wilting, discoloration, lethality etc. in intra- and inter-specific hybrid. Hybrid necrosis would result from an autoimmune-like response through epistatic interaction between resistance genes (Tonosaki *et al.*, 2016). Viewing heterosis as opposite to hybrid necrosis in plant growth suggests that, when necrosis is caused by induction of biotic stress genes, the repression of genes related to biotic stress may result in growth heterosis (Fujimoto *et al.*, 2018).

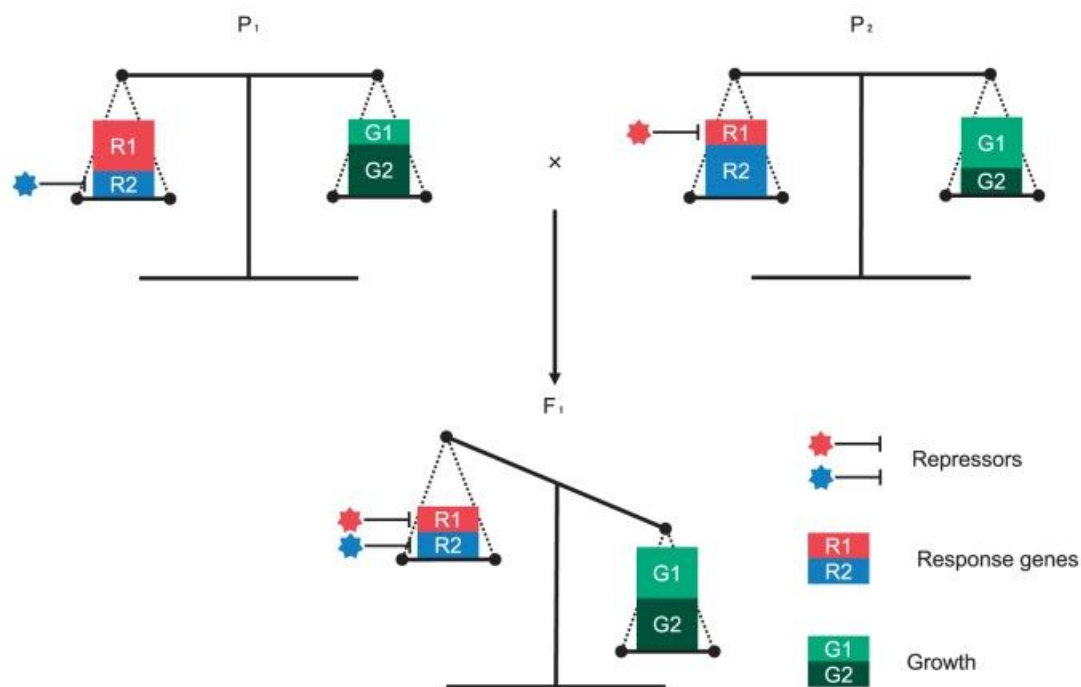
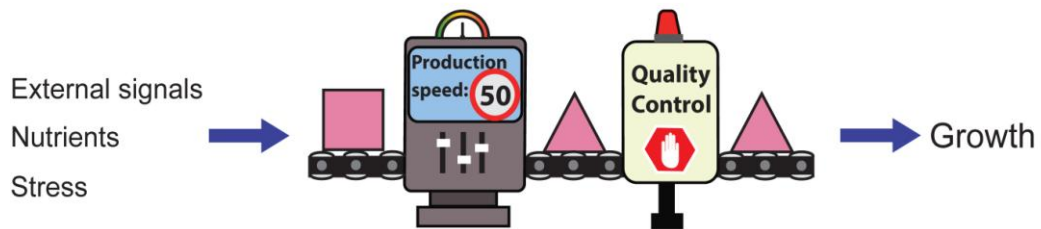


Figure 1.3: Model of trade-off between hybrid vigor and stress response gene expression proposed by Fujimoto *et al.* (2018).

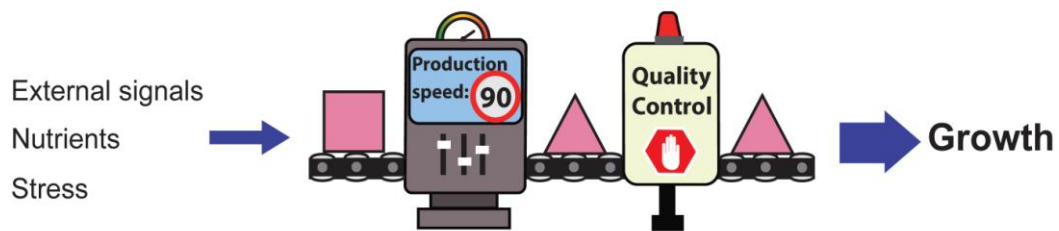
The model is further supported by studies showing, that genes involved in biotic stress in F_1 hybrids with growth vigor tend to be down-regulated compared to those in their parental lines (Wang *et al.*, 2008, Wang *et al.*, 2015). The repression of some defense response genes could therefore promote growth heterosis (Fujimoto *et al.*, 2018).

Bar-Zvi *et al.* (2017) discussed further two theories not mutually exclusive from the ones mentioned before (**Figure 1.4**). They are addressing the question, whether heterosis is a phenomenon enhancing an individual or an actually impairing effect caused by the clash of two genomes. The first theory models the widely believed enhanced hybrid growth through growth promoting factors. This is supported by studies which demonstrated the emergence of new favorable interactions between parental alleles in the hybrid. An example for this is SFT, where heterozygous combination increased yield by 60 % (Krieger *et al.*, 2010). Also, the dominant effect on the transcription factor AGAMOUS-LIKE 50 (AGL50), which was found in *Arabidopsis thaliana* (Seymour *et al.*, 2016), supports the idea of repairing recessive deleterious mutations in the hybrid, which accumulate in the inbred during inbreeding.

Parents:



Heterosis model class I: Enhancement of Growth Pathways



Heterosis model class II: Impairment of Growth Inhibition

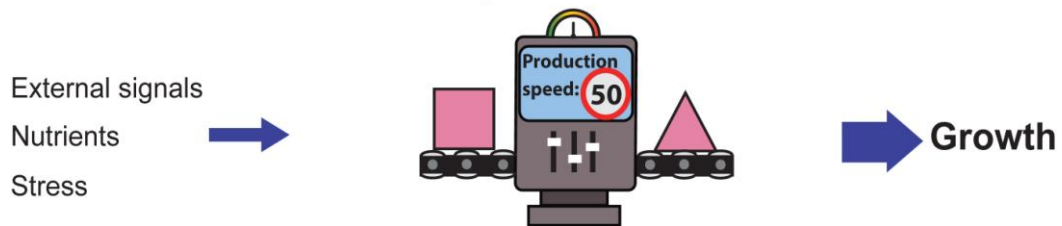


Figure 1.4: Model of hybrid vigor proposed by Bar-Zvi *et al.* (2017).

Heterosis models are divided into two classes; I. Enhanced hybrid growth is stimulated through enhancement or repair of growth promoting pathways. II. Enhanced hybrid growth is induced through the impairment of growth limiting pathways.

The second theory favors rather impairment (a deviation from the evolutionary optimum) as improvement (**Figure 1.4**). Molecular large-scale rewiring as evident in the hybrid may be more consistent with incompatibilities and dysregulation due to genome clashes than with improved performance (Bar-Zvi *et al.*, 2017). This is supported by findings of impaired growth-regulatory pathways, thus an inhibited size homeostasis (Herbst *et al.*, 2017). Reduction of transposable element (TE) regulation, which can lead to genome instability but activate gene expression, or transcription regulation present in hybrids (Freeling *et al.*, 2012) might explain the rare stabilization of the hybrid state (Bar-Zvi *et al.*, 2017).

1.3 Epigenetic mechanisms

Several studies provide evidence that changes of epigenetic factors, such as DNA methylation or chromatin modifications, can cause phenotypic variation and contribute to adaptation in plants (Kawakatsu *et al.*, 2016; Alakärppä *et al.*, 2018). In plants, DNA methylation, which is the transfer of a methyl group to a cytosine at its C5 position to form a 5-methylcytosine, is usually associated with transposon silencing (Law & Jacobsen, 2010), hence it is mostly found in repetitive regions of the genome (Teixeira & Colot, 2010). It is also known as gene regulator (Law & Jacobsen, 2010), for example in imprinting, which describes the parental specific expression of a gen, where the paternal and maternal alleles are differentially methylated (Scholten, 2010).

The loss of silencing through DNA methylation of one single locus, *FWA* (*flowering locus wageningen*) can lead to an early flowering phenotype (Soppe *et al.*, 2000). Moreover, spontaneous epimutations possibly linked to phenotypic variation in generations and among populations occur on a large range, and even more frequently than spontaneous genetic mutations, suggesting that sequence-independent epialleles play important roles in phenotypic diversity (Becker *et al.*, 2011; Kawakatsu *et al.*, 2016).

DNA methylation can be established on three different cytosine contexts in the plant genome: the symmetric CG and CHG, and the asymmetric CHH context (where H can stand for A, C or T, **Figure 1.5**).

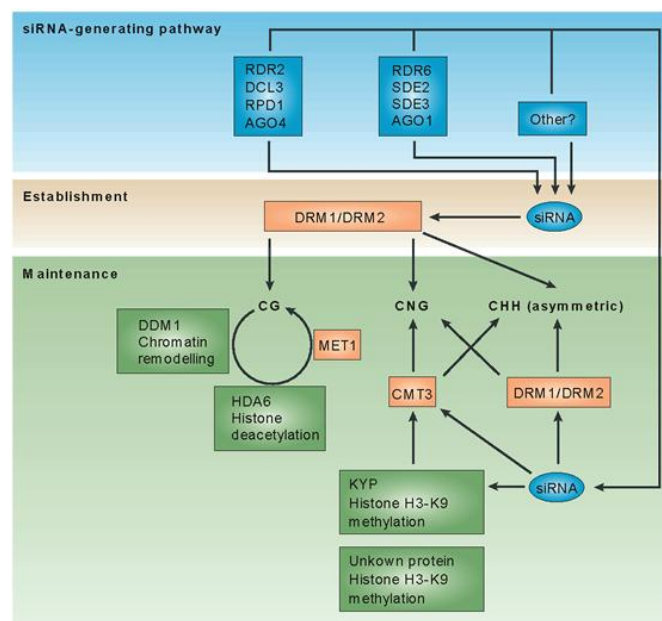


Figure 1.5: Methylation pathways of the symmetric (CG and CNG, where N is similar to H) and asymmetric (CHH) contexts in *Arabidopsis thaliana* (Chan *et al.*, 2005).

The methylation state of the symmetric contexts CG and CHG can be maintained during replication (**Figure 1.6**): CG context methylation is maintained by METHYLTRANSFERASE 1 (MET1) and DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), and non-CG contexts are maintained by DRM2, CHROMOMETHYLASE 2 (CMT2, Stroud *et al.*, 2014) and CMT3 (Bartee *et al.* 2001, **Figure 1.5**). In addition to the DNA methyltransferases, a chromatin remodeling factor, the ATP-dependent DNA helicase DDM1, is involved in the maintenance of DNA methylation, as described for example in *Arabidopsis* (Jeddeloh *et al.*, 1999) and rice (Higo *et al.* 2012, **Figure 1.6**).

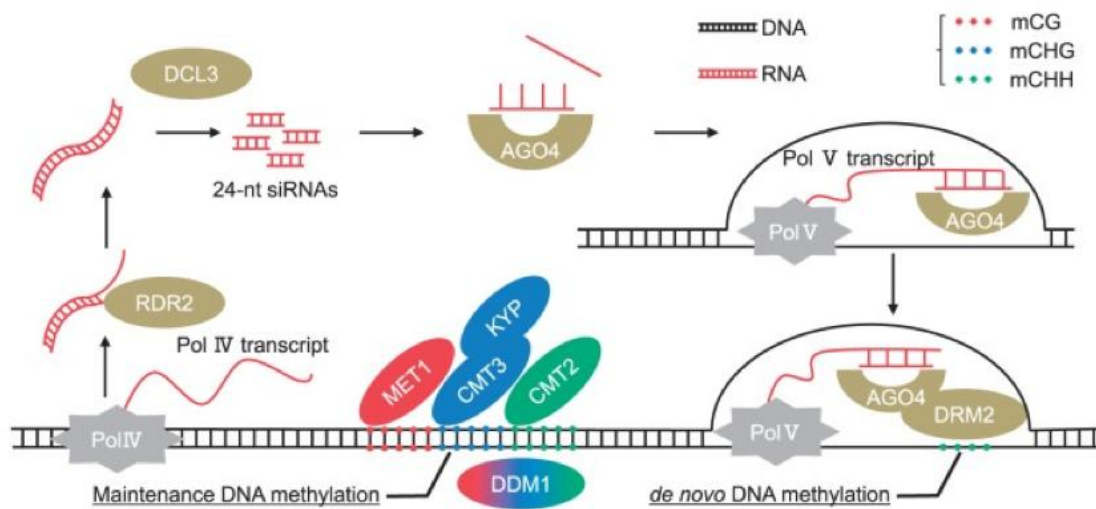


Figure 1.6: The maintenance methylation and RdDM pathway in plants (Fujimoto *et al.*, 2018).

All three contexts can be methylated *de novo* via RNA dependent DNA methylation (RdDM, **Figure 1.6**). This pathway is especially important for methylation of the asymmetric CHH context, but is also methylating the symmetric contexts. During RdDM, double-stranded RNAs are processed to 21-24 nt small interfering RNAs (siRNAs) by an enzyme cascade: Long single-stranded RNAs (ssRNAs) are produced by the transcription of repetitive regions via Polymerase IV (PolIV). The ssRNAs are converted to dsRNAs by RNA-dependent RNA polymerase 2 (RDR2) and cleaved by DICER-LIKE 3 (DCL3) before they are loaded onto the Argonaute protein AGO4 (Bologna & Voinnet, 2014). The siRNAs bind to scaffold RNAs transcribed by Polymerase V (PolV) at target loci, and therefore guide methylation via DRM2 (Lewsey *et al.*, 2015).

1.4 Effects of DNA methylation on heterosis

Due to its high regulatory impact, DNA methylation has a significant potential to exert effects on heterosis. Different studies imply that epigenetic changes in F1 hybrids are causally related to the heterotic effect (Shen *et al.*, 2012; He *et al.*, 2013). In early developmental stages, a highly dynamic DNA methylation patterning was found. Jullien *et al.* (2012) suggests interactions between different contributing parental (epi)genomes during initial post-fertilization development may establish the new states observed in heterotic hybrids.

In maize, DNA methylation differences between several genotypes (Eichten *et al.*, 2011, 2013) and between parental lines and hybrids (Liu *et al.*, 2014) indicate an epigenetic influence on the heterotic response. In *Arabidopsis*, an increased CG methylation was found in protein-coding as well as in transposon-coding regions in hybrids relative to their parental inbred lines (Greaves *et al.*, 2012; Shen *et al.*, 2012). The down-regulated genes include the circadian clock genes CCA1 and LHY, which have been shown to be involved in heterosis (Ni *et al.*, 2009). CHG methylation was observed to be slightly decreased in F1 hybrids (Greaves *et al.*, 2012). In both studies changes occur most frequently at loci where parental methylation levels differ.

In *Brassica napus*, Shen *et al.* (2017) found increased methylation levels at all three cytosine contexts in hybrids compared to their parents. The number of unmethylated cytosines was decreased, mainly those associated to genes, while the number of highly methylated cytosines was not increased. When analyzing differentially expressed genes (DEGs), the authors found an up-regulation of plant defense repressor expression and therefore a down-regulation in the salicylic acid pathway (Shen *et al.*, 2017).

In pigeonpea, the hypermethylation of CG and CHG contexts was associated to transposon-rich regions and CHH hypermethylation and was rather found in coding regions. However, the overall methylation levels exhibited an additive pattern in F1 hybrids (Junaid *et al.*, 2018). In contrast, lowering the methylation levels in inbred lines reduced the negative effect of inbreeding depression (Vergeer *et al.*, 2012). Liu *et al.* (2014) studied seedlings, embryos and endosperm of maize hybrids and their parental lines and showed lower relative total methylation levels in the hybrids than their corresponding mid-parental values.

Epigenetic recombinant inbred lines (epiRILs) with identical genetic, but different epigenetic background provide a powerful tool to analyze hybridization effects on an epi-

genetic level. Lauss *et al.* (2018) generated *Arabidopsis* epiHybrids by crossing Col wildtype maternal plants to *dmm-2-1* (DECREASE IN DNA METHYLATION1) derived epiRILs. They found high-parent heterosis for different quantitative traits (e.g. leaf area and plant height) induced by the decreased methylation rates in all three cytosine contexts which a reduced level of DDM1 is mediating. Dapp *et al.* (2015) observed heterosis for biomass in one of the epiHybrids using a similar approach by crossing a *met1*-derived epiRIL with Col wildtype.

Differences in expression levels of siRNAs between inbred parents and their hybrids indicated an influence of the hybrid state on methylation levels in various plant species (He *et al.*, 2010; Groszmann *et al.*, 2011; Barber *et al.*, 2012). Correlations between regions of altered methylation in general (Chodavarapu *et al.*, 2012) or CG methylation (Shen *et al.*, 2017) with changes in siRNA expression levels were observed. In general, a down-regulation of siRNAs in hybrids at least to the low parent level was observed (He *et al.*, 2010; Groszmann *et al.*, 2011; Barber *et al.*, 2012; Li *et al.*, 2012; Shen *et al.*, 2012). 24 nt siRNAs have the potential to act in *trans*, thus the sRNA derived from one parental allele can modify the epigenetic state of the other (Groszmann *et al.*, 2013; Greaves *et al.*, 2015). This results in the potential to generate F1 specific epigenomes with non-additive DNA methylation states caused by trans-chromosomal methylation (TCM), an increase in methylation at a locus with a low methylated allele from one parent gaining methylation to resemble the more heavily methylated allele of the other parent, or trans-chromosomal demethylation (TCdM), the loss of methylation at a loci adapting the low parent state). In *Arabidopsis*, increased DNA methylation in CG and CHH contexts was observed in hybrids together with 24 nt siRNAs associated to TCM and TCdM regions (Greaves *et al.*, 2012). Although TCM and TCdM events in hybrids depend on 24 nt siRNAs, mutations in PolIV and PolV disturbing RdDM did not affect the heterosis phenotype as observed by Zhang *et al.* (2016).

Additionally, parental sRNA variation (in contrast to genetic and expression variation) was discovered to be negatively associated with grain yield heterosis in maize. These sRNAs varying between parents were associated either to low or high grain yield, where the sRNAs associated to low heterosis were found to constitute a greater proportion (Seifert *et al.*, 2018).

1.5 Reduced Representation Bisulfite Sequencing

Cytosine methylation in plant genomes is important for a number of regulatory pathways. Genome-wide profiling of plant methylomes can be used to study phenotypic variations (Kwakatsu *et al.*, 2016), adaptation to environmental stresses or during development (Zhang *et al.*, 2010). However, in basic research as well as breeding programs, there is a need to monitor multiple samples to determine transgenerational methylation patterns or differential methylation. Methylome data obtained by bisulfite conversion followed by next-generation sequencing (NGS) provide genome-wide information on cytosine methylation (Schmidt *et al.*, 2017). Bisulfite sequencing (BS-seq) enables the detection of DNA methylation at cytosine residues at single-nucleotide resolution (Frommer *et al.*, 1992). Bisulfite sequencing is the treatment of genomic DNA with bisulfite that converts cytosine to uracil while leaving methylated cytosines unaffected (**Figure 1.7**).

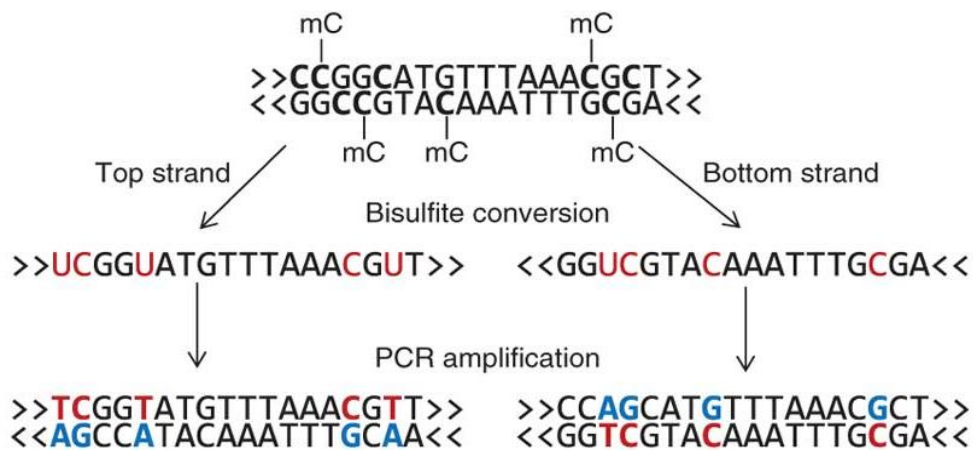


Figure 1.7: Principle of conversion of unmethylated cytosines to uracil through bisulfite (HSO_3^-) treatment. Unmethylated cytosines are deaminated to uracil, methylated cytosines are not converted (Krueger *et al.*, 2012).

However, the profiling of large populations or genomes is usually limited by the high costs for library preparation and sequencing. Reduced Representation Bisulfite Sequencing (RRBS) provides a cost-effective alternative compared to whole genome BS-seq, as it allows comparison of the selected regions over a range of samples with less need of sequencing depth (Bock *et al.*, 2010).

RRBS samples are generated by digesting genomic DNA with restriction endonucleases, e.g. *MspI*, which can be used for human DNA (Harris *et al.*, 2010). Other restriction en-

zymes like Taq^αI or ApeKI were used in studies to reduce the bias in coverage of genomic regions of different CpG densities (Lee *et al.*, 2014) or in a double enzyme digestion approach to increase accuracy of average methylation level detection (Wang *et al.*, 2013). The DNA restriction is followed by gel size selection and bisulfite conversion (Gu *et al.*, 2011). For the human or mouse genome, fragment size is selected in a range between 40 and 220bp in length, because this size has been shown to be plentiful in the sample and yield information on the vast majority of CpG islands (Andrews, 2013). DNA is fragmented and sequencing adaptors are ligated before bisulfite conversion, which may result in a loss of ligated DNA fragments due to DNA degradation by the bisulfite treatment. Minimizing DNA loss can be achieved by post-bisulfite adaptor tagging methods (Miura *et al.*, 2012; Smallwood *et al.*, 2014), where the bisulfite treatment is performed first.

1.6 Objectives

Heterosis or hybrid vigor offers a gain in crop improvement, e.g. in vitality and yield (Siddiq *et al.*, 2004; Fu *et al.*, 2014). Despite thorough research, the molecular basis of heterosis is still not completely defined. It is widely accepted that heterosis results from the combined action of diverged genomes and that genetic and epigenetic mechanisms are involved. Exploring heterosis on the epigenetic level will be of high value to find the contribution of DNA methylation to basic mechanisms leading to the formation of heterosis and provide useful information for application in hybrid breeding and cultivation.

Many studies in various plant species uncovered differences between DNA methylation levels of inbred lines and their hybrids. This study aims to analyze the contribution of DNA methylation to heterosis formation in maize and oilseed rape.

A RRBS protocol suiting the genomes of maize and oilseed rape and sample sizes will be adapted from existing protocols (Gu *et al.*, 2011; Smallwood *et al.*, 2014).

To analyze the effect of DNA methylation changes during early plant development on heterosis, maize early embryos (one day after pollination) treated with the demethylating agent 5-aza-2'-deoxycytidine (aza) are grown in an *in vitro* culture system. Seedling growth rate is measured as heterotic trait and compared between the inbred lines and their hybrid. RRBS is performed to confirm demethylation of the aza-treated embryo DNA.

To determine the significance of DNA methylation on heterosis formation and hybrid performance in a subset of a *Brassica napus* population, field data as well as *in vitro* root and shoot phenotype data of seedlings will be related to DNA methylation data generated by RRBS. DNA methylation, sRNA and mRNA expression patterns will be compared between hybrids and their parental lines and correlated to the phenotypic traits. Differential methylation between parents will be associated to a representative heterotic trait (root growth rate mid-parent heterosis).

2 Material and Methods

2.1 Materials

2.1.1 Plant material

a. Maize *in vitro* tissue culture

For the *in vitro* culture, inbred lines (A188, Ames22443, and H99, PI587129) were used grown under following controlled standard conditions in the greenhouse: Automatic irrigation, light intensity of max. 600 kLux/day and long-day light exposure (16 h day/8 h night). Controlled crossing was ensured through covering the ears with paper bags before the silks emerged. Pollination of A188 was performed directly on the cob at the plant with fresh pollen of either H99 or another A188 plant.

b. DNA methylation analysis in a rapeseed population

Seed material of oilseed rape (*Brassica napus* L.) double haploid lines, MSL (Male Sterility Lembke®) lines and hybrids was obtained from NPZ Innovation GmbH (Hohenlieth, Germany) and Deutsche Saatveredelung AG (Lippstadt, Germany).

The 471 lines contain 216 double haploid and 4 MSL lines from DSV, 204 double haploid lines, 10 MSL lines and 37 hybrids from NPZ. A subset („core factorial“) was selected from this breeding population containing all the hybrids with their respective parental lines (**Table 2.1**).

Table 2.1: Breeding scheme of the rapeseed population.

Paternal inbred lines are marked blue, maternal lines in red. Hybrid lines are marked in green.

		maternal lines						
		202	199	200	455	456	457	458
paternal lines	9			436		439		
	11			433		468		
	17			434		469		
	18			435		470		
	25			437		440		
	29		427	431		471		
	30				438			
	126		429					
	132		430					
	135						450	
	136							453
	454		428	432				
	459						445	
	460							451
	461						446	
	462	441						
	463						447	
	464	442					448	
	465	443					449	
	466	444						
467							452	

2.1.2 Chemicals, solutions and media

All used chemicals came, if not specified, from AppliChem GmbH (Darmstadt, Germany), Sigma-Aldrich Chemie GmbH (Munich, Germany) and Carl Roth GmbH & Co. KG (Karlsruhe, Germany). All enzymes and belonging buffers were, if not specified, obtained from New England Biolabs Inc. (Ipswich, USA). All prepared buffers and solutions were prepared with deionized water (MilliQ, Merck). Commonly used buffers and media were prepared, if not specified, according to the protocols of Sambrook *et al.*

(1989). The Murashige-Skoog medium including vitamins was obtained from Duchefa Biochemie B.V. (Haarlem, The Netherlands).

Table 2.2: Prepared solutions for *in vitro* culture.

Solution	Amount	Molecular formula	Name
BAP (2.5 mg/ml)	5 mg	$C_{12}H_{11}N_5$	6-Benzylaminopurine
			add. 2 ml EtOH

Thiamine HCl (2 mg/ml)	4 mg	$C_{12}H_{18}Cl_2N_4OS$	Thiamine Hydrochloride
			add. 2 ml ddH ₂ O

5-Aza-2'- deoxycytidine (2 mM)	4.5 mg	$C_8H_{12}N_4O_4$	5-Aza-2'-deoxycytidine
			add. 10 ml ddH ₂ O

Table 2.3: Solutions for gDNA extraction.

Solution	Amount	Molecular formula	Name
gDNA extraction buffer	6.68 ml	$C_{15}H_{28}NO_3$	30% N-Laurylsarcosine
	20 ml	$C_4H_{12}ClNO_3$	1M tris(hydroxymethyl)-methylammonium chloride
	4 ml	$C_{10}H_{16}N_2O_8$	0.5M Ethylenediamine-tetraacetic acid
	4 ml	NaCl	5M Sodium chloride
			add. 200 ml ddH ₂ O → pH 8.5 → filter sterile

Table 2.4: Prepared media for *in vitro* culture.

Solution	Amount	Molecular formula	Name
Murashige-Skoog medium (2x)	40 g	C ₁₂ H ₂₂ O ₁₁	Sucrose
	8.81 g		Murashige-Skoog medium including vitamins
			add. 1 l ddH ₂ O → pH 5.8 → filter sterile

modified Murashige-Skoog medium (2x)	250 ml		2x MS medium
	65 g	C ₁₂ H ₂₂ O ₁₁	Sucrose
	100 µl	C ₁₂ H ₁₈ Cl ₂ N ₄ OS	Thiamine Hydrochloride
	20 µl	C ₄ H ₈ N ₂ O ₃ *H ₂ O	6-Benzylaminopurine
	0.2 g	C ₄ H ₈ N ₂ O ₃ *H ₂ O	L-Asparagine monohydrate
			→ pH 5.8 → filter sterile

2.1.3 Used oligonucleotides

All oligonucleotides were obtained from biomers.net GmbH (Ulm, Germany) in HPLC grade. Primers and adapters were prepared in a 100 µM stock and further diluted for working solutions to 10 µM. Table 2.5 contains all used oligonucleotide sequences (* indicates phosphorothioate bonds; the underlined sequence indicates the eight-nucleotide index sequence) according to Quail *et al.* (2011) and Smallwood *et al.* (2014).

Table 2.5: Oligonucleotide sequences.

* indicates phosphorothioate bonds; the underlined sequence indicates the eight-nucleotide index sequence).

Name	Sequence
Biotin-labeled Oligo1	5' -(Biotin)CTACACGACGCTCTTCCGATCTNNNNNNNNN- 3'

Oligo2	5' -TGCTGAACCGCTCTCCGATCTNNNNNNNNN- 3'
PE1.0 forward primer	5' -AATGATACGGCGACCACCGAGATCTACACTCTCCCTACACGACGCTCTCCGATC*T- 3'
iPCRtagT 1	5' -CAAGCAGAAGACGGCATAACGATGAGATCAACGTGATGAGATCG GTCTCGGCATTCTGCTGAACCGCTCTCCGATC*T - 3'
iPCRtagT 2	5' - CAAGCAGAAGACGGCATAACGATGAGATCAACATCGGAGATCGGTCTCGGCATTCTGCTGAACCGCTCT TCCGATC*T - 3'
iPCRtagT 3	5' - CAAGCAGAAGACGGCATAACGATGAGATATGCCTAAGAGATCGGTCTCGGCATTCTGCTGAACCGCTCT TCCGATC*T - 3'
iPCRtagT 4	5' - CAAGCAGAAGACGGCATAACGATGAGATAGTGGTCAGAGATCGGTCTCGGCATTCTGCTGAACCGCTCT TCCGATC*T - 3'
iPCRtagT 5	5' - CAAGCAGAAGACGGCATAACGATGAGATACCACTGTGAGATCGGTCTCGGCATTCTGCTGAACCGCTCT TCCGATC*T - 3'
iPCRtagT 6	5' - CAAGCAGAAGACGGCATAACGATGAGATACATTGGCGAGATCGGTCTCGGCATTCTGCTGAACCGCTCT TCCGATC*T - 3'
iPCRtagT 7	5' - CAAGCAGAAGACGGCATAACGATGAGATCAGATCTGGAGATCGGTCTCGGCATTCTGCTGAACCGCTCT TCCGATC*T - 3'
iPCRtagT 8	5' - CAAGCAGAAGACGGCATAACGATGAGATCATCAAGTGAAGATCGGTCTCGGCATTCTGCTGAACCGCTCT TCCGATC*T - 3'
iPCRtagT 9	5' - CAAGCAGAAGACGGCATAACGATGAGATCGCTGATCGAGATCGGTCTCGGCATTCTGCTGAACCGCTCT TCCGATC*T - 3'
iPCRtagT 10	5' - CAAGCAGAAGACGGCATAACGATGAGATACAAGCTAGAGATCGGTCTCGGCATTCTGCTGAACCGCTCT TCCGATC*T - 3'

2.1.4 Software

Table 2.6: Software used in this study.

Name	Version	Author/Company	Function	Download source
Bismark	0.18.1	Krueger & Andrews (2011)	Mapping of bisulfite converted sequence reads and determination of cytosine methylation state	https://www.bioinformatics.babraham.ac.uk/projects/bismark/
BLASTx	2.2.26+	Camacho <i>et al.</i> (2009)	sequence alignment	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
Bowtie2	2.3.2	Langmead & Salzberg (2012)	sequence mapping	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
CGmaptools	0.1.1	Guo <i>et al.</i> (2018)	Methylation data analysis (DMR calling)	https://github.com/guoweilong/cgmaptools
Cutadapt	1.5	Martin (2011)	quality and adapter trimming of fastQ files	https://github.com/marcelm/cutadapt
FastQC	0.11.4	Andrews (2010)	quality analysis of FastQ files	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
methyKit	1.0.0	Akalin <i>et al.</i> (2012)	R package for methylation data analysis (DMP calling)	https://bioconductor.org/packages/release/bioc/html/methyKit.html
Python2/3		Python Software Foundation, DE, USA	programming language	https://www.python.org/downloads/
R		Ihaka & Gentleman (1996)	statistical analysis and plotting	http://www.r-project.org/
SAMtools	1.3	Li <i>et al.</i> (2009)	SAM/BAM file handling	http://samtools.sourceforge.net
Trimgalore	0.4.2	Krueger (2015)	quality and adapter trimming of Fastq files	https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

2.1.5 Data

2.1.5.1 Phenotypes of the rapeseed population

Field trait measurements were taken from the collaborating breeding companies NPZi and DSV from several independent field trials (**Table A.1**). Measured traits were yield [dt/ha], plant length [cm], thousand kernel weight [mg], field emergence [d], oil content [%], glucosinolate content [%] and protein contents in seeds [%]. The measurements were taken in two consecutive years from five different locations, combined and normalized by collaboration partners from the department of plant breeding at the Justus-Liebig-University Gießen.

Root and shoot measurements were taken *in vitro* with a phenotyping platform by the Jülich Plant Phenotyping group (**Table A.2**). The following features were measured: Total root length [cm], taproot length [cm], lateral root length [cm], root system depth [cm], root system width [cm], convex hull area [cm²], lateral root branching angle [cm], number of laterals, shoot dry weight [mg], root dry weight [mg] and root-to-shoot ratio. The length measurements were taken on the fourth, eighth, 14th and 18th day after sowing (das).

2.1.5.2 RNA library sequences of rapeseed population

For small RNA and mRNA library preparation, total RNA was extracted from pooled seedlings of *Brassica napus* (2.2.1.2) using the Quick-RNA Miniprep kit (Zymo Research Europe GmbH, Freiburg, Germany). Quality and quantity was checked on a fragment analyzer using the Standard sensitivity RNA kit (Advanced Analytical Inc., Ankeny, USA). Only RNA with a RQN (RNA quality number) ≥ 6.5 was included for sequencing library preparation. Small RNA libraries were prepared by the technician Dominika Rybka using the NEBNext Multiplex Small RNA library prep set for Illumina (New England Biolabs Inc., Ipswich, USA) according to the manufacturer's protocol. Pooled libraries per lane were sequenced at BGI in Hong Kong (BGI Genomics Co., Ltd., Shenzhen, China) with approximately 10 million 50 bp single-end reads on a HiSeq4000 (Illumina Inc., San Diego, CA, USA).

Messenger RNA libraries were prepared by the technician Dominika Rybka using the NEBNext Ultra Directional RNA library prep kit for Illumina (New England Biolabs Inc., Ipswich, USA) according to the manufacturer's protocol. Pooled libraries were sequenced at BGI in Hong Kong (BGI Genomics Co., Ltd., Shenzhen, China) with approxi-

mately 10 million 100 bp paired-end reads on a HiSeq4000 (Illumina Inc., San Diego, CA, USA).

2.1.5.3 Processed RNA reads

For small RNA raw sequencing data, reads were processed and normalized by Dr. Felix Seifert resulting in read counts per million quantile normalized (rpmqn) according to Seifert *et al.* (2018).

For mRNA raw sequencing data, reads were processed by the Christian Rockmann from the Technische Hochschule Wildau. Mapping was performed using HISAT2 (version 2.1.0, <https://ccb.jhu.edu/software/hisat/index.shtml>) under default conditions and --dta option (creating some additional flags for downstream transcriptome assembly). Assembly of RNA-Seq alignments into potential transcripts was performed using StringTie (version 1.3.4d, <https://ccb.jhu.edu/software/stringtie/#install>) under default conditions. Differential expression of transcripts was called by using CuffDiff (version 2.2.1, <http://cole-trapnell-lab.github.io/cufflinks/cuffdiff/>). Only significantly differential fold changes of transcripts were included in further analysis.

2.2 Methods

2.2.1 Plant growth and cultivation

2.2.1.1 Maize tissue culture

For the early embryonic maize tissue culture, a modification of the embryo sac culture method of Campenot *et al.* (1992) was used. One day after pollination fertilized embryo sacs with nucellus tissue were collected on aza-containing (50 μ M) or control modified MS medium (**Table 2.4**) with the cut surface facing the medium (Campenot *et al.*, 1992) and cultured in the dark at 26 °C to 28 °C three to six weeks (**Figure 2.1 A-B**). After about 6 d (depending on the developmental stage), a portion of the embryos was isolated for bisulfite sequencing, and the rest were cultured until germination (3-7 weeks). Two days after germination (dag, **Figure 2.1 C**), the germ buds were transferred to standard MS media (Murashige & Skoog, 1962) and stored in the phytochamber (24 °C/16 h). Starting six days after transfer, the plant height was measured on every second day for a two-week period to determine seedling growth rate in their linear growth phase (**Figure 2.1 D**) from 8 dag and 22 dag. Growth heterosis was

calculated via quotient from individual growth rate mean of hybrids and general growth rate mean of inbred lines for Better Parent Heterosis (BPH, formula 1) and Mid-Parent Heterosis (MPH, formula 2).

Formula 1: $\frac{F_1}{BP}$, where F_1 is the growth rate value of every F_1 hybrid and BP is the mean growth rate value of the better parent.

Formula 2: $\frac{F_1}{MP}$, where F_1 is the growth rate value of every F_1 hybrid and MP is the mean growth rate value of the two parents. For statistical analysis the two-tailed, homoscedastic Student's t test (two-sample equal variance) was performed with growth rate values with $R^2 > 0.90$ only. The better parent A188 was used as maternal line for extensive aza tests due to much higher regeneration rates compared to H99.

For analyzing the DNA demethylation effects of the aza treatment, 6 developing embryos were isolated 7 days after pollination (dap) from the surrounding nucellus tissue from either aza containing or control plates, flash frozen in mannitol and pooled for DNA extraction. Seedlings were sampled after the last growth measurement at 22 dag, cleaned from the medium with a water flush and flash frozen. The entire seedling was used for DNA extraction to investigate the continuity of DNA demethylation.

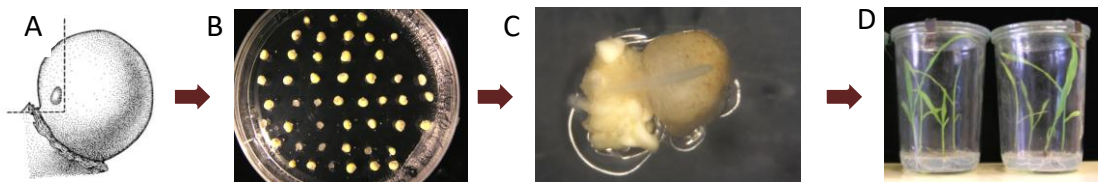


Figure 2.1: Process of the nucellus in vitro culture.

A. The embryo is dissected with the surrounding nucellus tissue 1 dap. **B.** Embryos are regenerating after 3 to 4 weeks on either aza-containing medium or control medium. **C.** Germination takes place c. after one month and germinating buds are transferred to MS medium without aza. **D.** Seedlings plant height measurements were performed 6 days after transfer and were taken for the following two weeks.

2.2.1.2 Cultivation of the rapeseed population

The rapeseed seeds (2.1.1) were sown in a randomized sowing scheme in 150 pot planting trays filled with potting soil with five biological replicates per line (**Figure 2.2 A**). The plants were germinated and grown in a growth chamber under day/night temperatures of 24/18 °C and a 16 h photoperiod. The planting trays were moved in the growth chamber every following day to provide similar growing conditions to all plants. Plant material was collected when the first primary leaf was fully unfolded (c.

10-13 days after sowing, **Figure 2.2 B**). Whole seedlings including roots were cleaned from soil, flash frozen in liquid nitrogen and stored at -70°C . Five biological replicates of each line were pooled in liquid nitrogen, ground to fine powder and stored at -70°C .

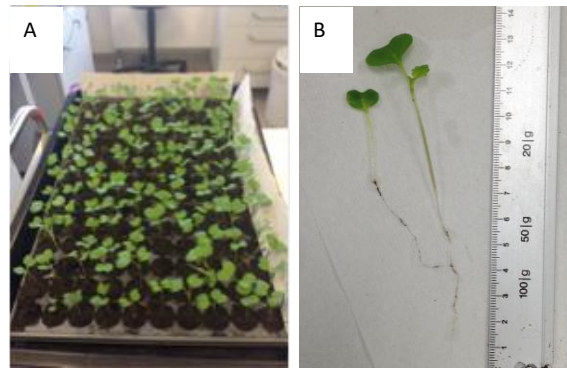


Figure 2.2: Seedlings of *Brassica napus* genotypes at time of sampling.

A. Whole tray with the seedlings at randomized positions. **B.** Two different genotypes during measurement.

2.2.2 Reduced Representation Bisulfite Sequencing (RRBS)

This method can be used to analyze defined fractions of genomes from limited samples by the application of restriction digestion, gel separation and fragment elution prior to BS-seq library preparation. The reduction of represented genomic regions lowers the sequencing cost considerably while providing an adequate sample of genome-wide DNA methylation pattern.

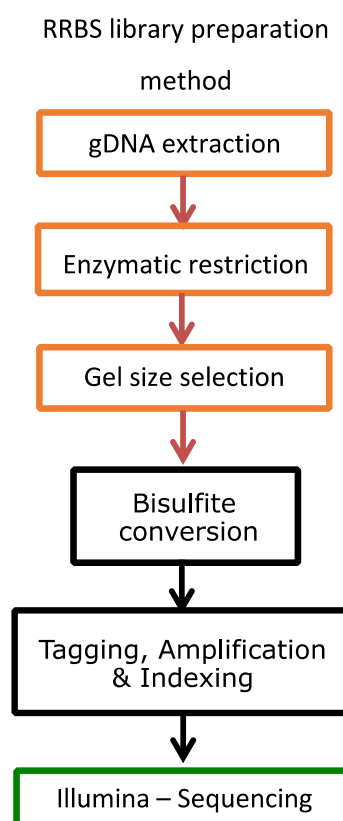


Figure 2.3: Schematic overview over the RRBS library preparation method.

After gDNA extraction, DNA restriction with a methylation-insensitive enzyme is performed. Fragments of appropriate size are selected by common agarose gel electrophoresis and a custom elution protocol. This takes care for reduced representative genomic regions that are comparable over a bulk of samples.

The RRBS technique (**Figure 2.3**) involves the bisulfite conversion of unmethylated cytosines to uracil through deamination (which is the removal of an amino group of a molecule). During amplification, the uracil is transcribed to thymidine. The conversion ensures the detection of methylated cytosines (staying C) and unmethylated cytosines (which are visible as T) in the later mapping procedure (**Figure 1.7, 1.5**).

In this approach the bisulfite treatment is performed first, which lead to simultaneous DNA fragmentation and conversion of unmethylated cytosines to uracil. Complementary strand synthesis is primed by 5' biotin modified oligonucleotides containing Illumina-compatible adaptor sequences and a stretch of nine random nucleotides at the 3' end. After capturing the tagged strands with streptavidin-coated magnetic beads, a second adaptor is integrated in the same way, and PCR amplification is per-

formed with indexed primers. The converted and indexed libraries are purified and sequenced.

2.2.2.1 Optimization of RRBS protocols

a. gDNA digestion and size selection

To obtain the optimal fragment size distribution for subsequent gel size selection and bisulfite conversion, three different restriction enzymes and their combinations were tested (Wang *et al.*, 2013; Lee *et al.*, 2014): Taq^αI (restriction site TCGA), MspI (restriction site CCGG) and ApeKI (restriction site GCWGC).

Restriction assays were performed with 1 µg gDNA of the *Brassica napus* cultivar Drakkar according to the enzyme's manufacturer's protocol with 10 U per reaction for 1 h with temperatures of 37 °C for MspI, 65 °C for Taq^αI and 75 °C for ApeKI. Double restrictions were performed with the either lower temperature first for 1 h followed by the digestion at the higher temperature for 1 h. Cleanup of all reactions was performed with magnetic beads (Agencourt AMPure XP beads, Beckman Coulter Inc., Indianapolis, USA) on a 1:1 ratio according to the manufacturer's guidelines. Quality and quantity were determined either by agarose gelelectrophoresis or on a Fragment Analyzer (Advanced Analytical Inc., Ankeny, USA).

A DNA fragment size of 1000 – 1200 bp was selected and cut from an EtBr-stained agarose gel and the DNA was eluted with a custom column consisting of folded Whatman filter paper (Sigma-Aldrich Chemie GmbH, Munich, Germany) in a 0.5 ml tube with a hole in the bottom placed in a 2 ml tube (**Figure 2.4**).

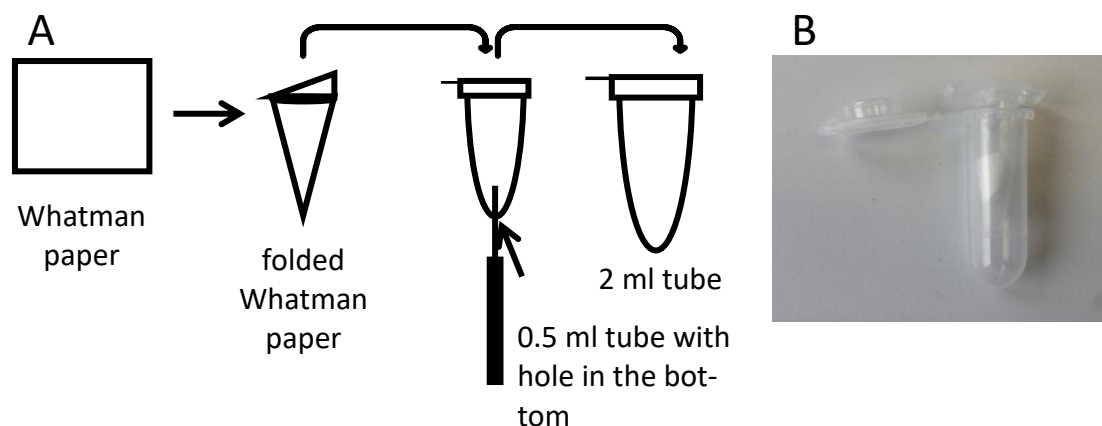


Figure 2.4: Structure of a custom column to elute DNA from an agarose gel.

A. The schematic overview shows the folded Whatman filter paper, which is located in a 0.5 ml tube with a subtle hole in the bottom. The 0.5 ml tube sits in a 2 ml tube to collect the dissolved DNA. **B.** Photograph of the custom column.

b. RRBS library preparation optimization

For comparison of two different bisulfite sequencing protocols (NextFlex Bisulfite Sequencing Kit (Bioo Scientific Corporation, Austin, USA) and custom method modified after Smallwood *et al.*, 2014) gDNA of ~300 mg ground plant material of one maternal line (5) and one paternal line (4) was isolated via phenol-chloroform extraction described by Pallotta *et al.*, (2000) with the following change: For protein degradation, 800 μ l of phenol-chloroform-isoamyl alcohol (25:24:1) was added to the sample tissue/ extraction buffer-mix once and shaken thoroughly for 5 min.

One of the two protocols tested was developed by Smallwood *et al.*, 2014 and adjusted to RRBS (Edelmann & Scholten, 2018). The method was performed as described in 2.2.2.2.

The NEXTflex Bisulfite-Seq Kit (5119-01, Bioo Scientific Corporation, Austin, USA) was tested as second RRBS library preparation method. Size selection, bisulfite conversion, ligation and amplification (18 cycles) were performed according to the manufacturer's protocol. Quality and quantity of all produced libraries were tested on a Fragment Analyzer (Advanced Analytical Inc., Ankeny, USA).

The sequencing libraries of both RRBS protocols were sequenced by BGI Tech Solutions Co., Ltd. (Hong Kong, China) with 50 bp SE reads on a HiSeq2000 (Illumina Inc., San Diego, CA, USA).

2.2.2.2 RRBS library preparation

Bisulfite sequencing library preparation of Smallwood *et al.* (2014) was optimized and adapted to RRBS (Gu *et al.*, 2011), which resulted in the following protocol (Edelmann & Scholten, 2018):

Genomic DNA from plant material (2.1.1, both maize and rapeseed) was isolated as described in 2.2.2.1.b. Extracted gDNA was digested with Taq^oI for 1.5 h at 65 °C. DNA fragment size selection was performed as described in 2.2.2.1.a. The one-step-modification-procedure of the Imprint DNA Modification Kit (Sigma-Aldrich, Chemie GmbH, Munich, Germany) was performed followed by column cleanup with the following variations: After denaturation at 99 °C and incubation at 65 °C, the converted DNA was stored over night at 7 °C. In the cleanup, the centrifugation steps were generally set to 30 s except for drying the column after the second ethanol wash step, which was set to 1 min, followed by a 3 min incubation at room temperature with an open lid to ensure complete ethanol removal. The elution solution was preheated to 60 °C and elution centrifugation was performed twice with 10 µl of elution buffer each.

After desulfonation, random priming was performed through mixing the 20 µl of converted DNA with 1 µl dNTPs (10 mM, Bioline GmbH, London, UK), 1 µl oligo 1 (100 µM, **Table 2.5**) and 2.5 µl Blue Buffer (10x, Biozym Scientific GmbH, Hessisch Oldendorf; Germany) before incubation at 65 °C for 3 min followed by 4 °C pause. 1 µl of high concentrated Klenow (50 U/µl, 3'-5' exo⁻, Biozym Scientific GmbH, Hessisch Oldendorf; Germany) was added and the samples incubated at 4 °C for 5 min. The incubation temperature was raised with +1 °C every 15 s up to 37 °C, followed by extension at 37 °C for 30 min. Subsequently, the samples were incubated with 3 µl exonuclease I (20 U/µl, New England Biolabs Inc., Ipswich, USA) for 1 h at 37 °C before purification using 0.8x HighPrep PCR magnetic beads (MagBio Genomics Inc., Gaithersburg, USA) according to the manufacturer's protocol. Samples were eluted in 40 µl 10 mM Tris-HCl (pH 8.5) and incubated with washed M-280 Streptavidin Dynabeads (Invitrogen by Fisher Scientific GmbH, Schwerte, Germany) for 20 min with rotation at room temperature. Beads were washed twice with 100 µl 0.1 N NaOH, and twice with 100 µl 10 mM Tris-HCl (pH 8.5) and resuspended in a mix of 41 µl ultrapure H₂O, 5 µl Blue Buffer

(10x), 2 μ l oligo 2 (100 μ M, **Table 2.5**) and 2 μ l dNTPs (10 mM). Samples were incubated at 95°C for 45 s and transferred immediately to ice before addition of 2 μ l Klenow exo- (50 U/ μ l) followed by incubation at 4 °C for 5 min, +1 °C/15 s to 37 °C, and extension 37 °C for 90 min. The beads were washed once with 100 μ l 10 mM Tris-HCl, pH 8.5) and resuspended in 21 μ l ultrapure H₂O, 25 μ l MyFi Mix (2x, Bioline GmbH, London, UK), 2 μ l PE1.0 forward primer (10 μ M, **Table 2.5**) and 2 μ l of indexed iPCRTag reverse primer (10 μ M, **Table 2.5**). PCR amplification of the libraries was performed as follows: 95 °C 2 min, 12 repeats of (95 °C 30 s, 65 °C 30 s, 72 °C 30 s). Final extension at 72 °C for 3 min was followed by 4 °C hold. Amplified libraries were purified using 0.8x HighPrep PCR magnetic beads according to the manufacturer's guidelines and quality and quantity assessment was performed using a Fragment Analyzer (Advanced Analytical Inc., Ankeny, USA).

Pools of about 37 libraries per lane were sequenced at BGI in Hong Kong (BGI Genomics Co., Ltd., Shenzhen, China) with approximately 10 million 100 bp paired-end reads on a HiSeq4000 (Illumina Inc., San Diego, CA, USA).

2.2.3 Computational methods

2.2.3.1 RRBS read preparation and analysis

The 50 bp sequencing data was trimmed off adapter sequences and low-quality sequencing results (phred score 20) and subsequently freed from adapter contamination (in-house developed programs). Reads generated with the custom RRBS adapted from Smallwood *et al.* (2014) protocol were trimmed of 9 bases at the 5'-end of each read originating from the random priming during adapter ligation to avoid any bias. Processed sequencing data was quality checked using FastQC (0.11.4) and reads not mapping the reference genome were mapped to the chloroplastial and mitochondrial genome (http://plants.ensembl.org/Brassica_napus/Info/Index (Kersey *et al.*, 2017), genome assembly AST_PRJE5043_v1) to test for successful bisulfite conversion. The conversion rate was calculated as percentage of methylated cytosine contexts in comparison to all sites. *Brassica napus* genome (<http://www.genoscope.cns.fr/brassicnapus/> (Chalhoub *et al.*, 2014), version 4.1) was prepared to suit bisulfite converted reads with the Bismark genome preparation module. The converted reads were mapped to this prepared genome via Bismark. Further analysis for testing was done with in-house written scripts.

The 100 bp paired end sequencing data was trimmed of adapter sequences and low-quality reads (phred score 20) using Trim galore (version 0.4.2). Additionally, 9 bases were removed from each end of the reads (as described above). Processed sequencing data was quality checked using FastQC (0.11.4). Bisulfite conversion rate was estimated via mapping of the unmapped reads to the bisulfite converted mitochondrial and chloroplastial *Zea mays* (<ftp://ftp.ensemblgenomes.org/pub/plants/>, version AGPv4) and *Brassica napus* (<http://www.genoscope.cns.fr/brassicanapus/>, version 4.1) genome via Bismark (version 0.18.1) and mapping to the respective reference genomes (<ftp://ftp.ensemblgenomes.org/pub/plants/>, version AGPv4; <http://www.genoscope.cns.fr/brassicanapus/>, version 4.1) was performed in the nondirectional mode (as recommended for library constructed in a non strand-specific manner) and with a changed Bowtie2 valid alignment score allowing inclusion reads with more mismatches and read gaps (--score_min L,0,-0,4,). Methylated cytosines were called via Bismark methylation extractor (version 0.18.1) with regard to the generated “CX-report” containing all information useful for subsequent analysis. A read coverage threshold of five reads per cytosine was used to ensure coverage flexibility and reflection of a real situation in later testing and analysis.

2.2.3.2 sRNA read preparation

Mapping of sRNAs on the reference genome of *Brassica napus* (<http://www.genoscope.cns.fr/brassicanapus/>, version 4.1) was performed using HISAT2 (Kim *et al.*, 2015) with no softclipping and mismatches allowed on all possible mapping loci (multiple mapping), to avoid the exclusion of the majority of sRNAs usually mapping on several loci due to their origin in repetitive regions. For length analysis, only one sRNA per index was included to avoid a distribution bias caused by multiple mappings.

2.2.3.3 Characterization of DNA methylation and correlation to phenotypes

For statistical analysis and plotting of the processed bisulfite converted reads, custom Python (version 3.6.4) and R (version 3.3.3) scripts were used utilizing implemented modules, build-ins and packages (**Table 2.6**). Differentially methylated positions (DMPs) were called using the R package methylKit (Akalın *et al.*, 2012). The methylKit DMP calling is based on Fisher’s exact test, when no replicates are present and on logistic regression in the presence of replicates, which in this case means the calculation of DMPs between the groups. DMP significance is given by *P*-values that are corrected

to q-values using the sliding linear model (SLIM) method (Akalin *et al.*, 2012). MethyKit was used under default conditions, which exclude regions with q-values < 0.01%. Methylation differences of $\geq 25\%$ and $\geq 50\%$ for CpG, $\geq 10\%$ and $\geq 25\%$ for CHG and $\geq 2\%$ and $\geq 5\%$ depending on the overall methylation level of the contexts were tested for their DMP output amount. Differentially methylated regions (DMRs) were called using the CGmap software (Guo *et al.*, 2018), which is based on dynamic fragmentation strategy. Several thresholds determine length and position of each DMR: Only regions with at least 3 cytosines covered and a maximum distance of 100 bp between them were included. The maximum size of each fragment should not exceed 1000 bp. The software defines DMRs as regions with a significant minimum mean methylation difference of 10% ($p \leq 0.01$, *t*-test, Guo *et al.*, 2018). Additionally, only DMRs with a minimum methylation difference of 25% (Ziller *et al.*, 2015) were extracted from these for further analysis. DMPs and DMRs were called between the maternal and paternal parents of each hybrid and compared to phenotypic and methylation data of the hybrids, e.g. for inheritance pattern and TCM/ TCdM (trans-cromosomal methylation/transcromosomal demethylation) events.

For correlation of methylation and phenotypic data, field data and root data from collaboration partners belonging to the corresponding lines were used (2.1.1). Pearson correlation was performed using custom Python and R scripts. To finally correlate field traits, root and shoot growth in seedlings and the methylation state, the Pearson correlation, a measure of the linear correlation between two variables, was used (Pearson, 1895).

Combining methylation data or DMRs with mRNA or sRNA data (2.1.5.3) and annotated features by location (e.g. calculation of TCM or TCdM events as defined in Greaves *et al.*, 2012) was done using custom python scripts, where locations of 500 bp or 1000 bp, respectively, up- and downstream of each feature were included, and using the publicly available *Brassica napus* annotation version 5 (<http://brassicadb.org/brad/index.php>) for coding sequences (CDS) and the corresponding repeat masker data TE.masking.gff (<http://www.genoscope.cns.fr/brassicanapus/data/>) for repetitive elements. Analysis of DNA methylation, mRNA and sRNA patterns at specific enzymes was performed using custom python scripts, which compare the location of sample and respective enzyme (<https://www.ncbi.nlm.nih.gov/gene>, search for *Brassica napus*) including

1000 bp up- and downstream (**Table A.3**). The mean was calculated for each location or feature, respectively.

Association of parental differentially methylated regions to MPH for total root length (TRL) of seedlings and yield (y) was used as developed by Seifert *et al.* (2018) for parental sRNA expression differences, by separating the hybrids into classes of low and high trait values with equal size. For each parental DMR the number of corresponding hybrids was counted for both classes. The binomial distribution probability function was applied to generate p values (Frisch *et al.*, 2010), which were further corrected by Benjamini-Hochberg FDR correction (Benjamini & Hochberg, 1995). DMRs with $p \leq 0.05$ were considered as associated to MPH for TRL and termed heterosis-associated DMRs (ha-DMRs). Negatively heterosis-associated DMRs were defined as DMRs with significantly (as mentioned before in this paragraph) higher emergence in the low trait hybrid group, whereas positively ha-DMRs display significant appearance in the high performing hybrid group (Seifert *et al.*, 2018).

For statistical analysis using the Student's t-test, the two-tailed, homoscedastic Student's t-test (two-sample equal variance) was performed.

3 Results

3.1 Optimization of the RRBS protocol

The RRBS method was used to analyze fractions of genomes from limited samples by the application of restriction digestion, gel size selection and BS-seq library preparation including bisulfite conversion. Restriction and size selection ensures enrichment of similar genomic regions per sample. The reduction of represented regions lowers the sequencing cost considerably through less sequencing depth needed, while providing a sample of genome-wide DNA methylation pattern. Several variations of the different steps of the protocol were investigated to fit and adapt the RRBS method to the used organism gDNA. The optimizations of the subsequent steps are described in the following passages 3.1.1, 3.1.2.

3.1.1 gDNA digestion and size selection

For gDNA digestion, three different restriction enzymes and their combinations of two restriction enzymes were tested using gDNA of the *Brassica napus* cultivar Drakkar for fragment length distribution patterns (2.2.2.1).

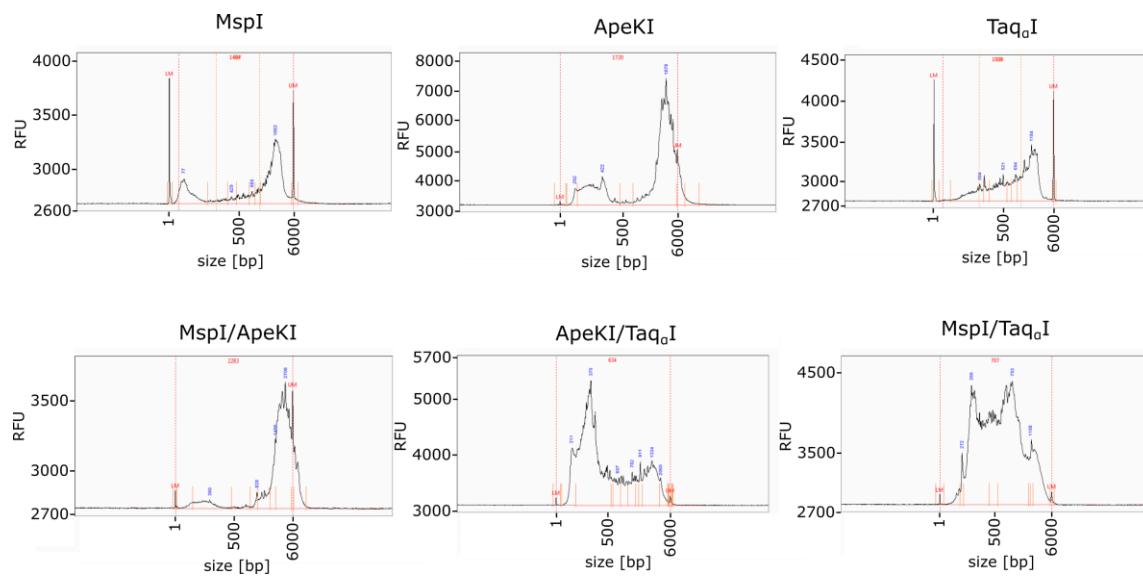


Figure 3.1: Digestion with different enzymes and their respective combinations.

The digestion revealed different fragment size distribution patterns per used restriction enzyme between 1 and 6000 bp length. The electropherograms given by the Fragment Analyzer after digestion with the single restrictions are shown in the upper part and the double digestions in the lower part of the figure.

Single digestion of gDNA with either MspI or ApeKI resulted in fragment sizes between 10 to 5000 bp and 6000 bp, respectively (**Figure 3.1**). Double digestions using the same enzymes showed similar size distributions. The single and double enzyme gDNA digestions including MspI revealed a less amount of fragments with a size of about 300 to 700 bp, whereas all gDNA restrictions that included ApeKI exhibited “valleys” in the range of 400 to 1000 bp. The gDNA digestions with Taq^qI resulted in an evenly distributed fragment length pattern with both fragment length and concentration increasing (**Figure 3.1**).

Gel size selection was performed to reduce the number of genomic regions present in the later sequencing libraries. This yielded in representative regions in all samples of the population, which ensures comparability after sequencing. **Figure 3.2** shows gDNA digested by Taq^qI before and after gel size selection. A fragment size of 1000 to 1200 bp was selected to compensate for further fragmentation through following bi-sulfite conversion.

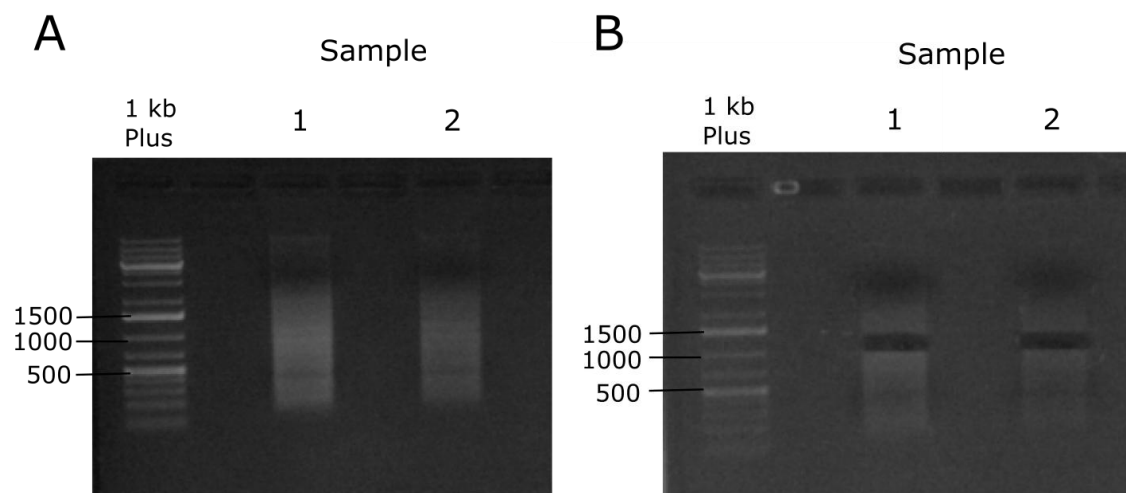


Figure 3.2: Gelectrophoretic separation of Taq^I restricted *Brassica napus* gDNA.

Separation was performed using a 1 % Ethidiumbromide/agarose gel of two different *Brassica napus* genotypes before (A) and after (B) gel size selection.

Estimation of potential cytosine coverage by *in silico* restriction and fragment size selection using reference genome sequences as templates confirmed the suitability of Taq^I as restriction enzyme for both maize and oilseed genome (Figure 3.3). The *in silico* restriction was performed by using a size of 1000-1200 bp (used for gel size selection) and 40-220 bp (fragment size used in other RRBS protocols and assumed to be a possible outcome after additional fragmentation by bisulfite conversion). Whereas all digestions enzymes produced similar fragment sizes when using a size selection of 100-1200 bp in maize and rapeseed, size selection of 40-220 bp differed in fragment size distribution outcome per enzyme and species. The highest percentage of genome and cytosine coverage was achieved by *in silico* digestion with Taq^I for both, the maize (13 %, 14 %) and rapeseed (9 %, 12 %) reference genome. Both MspI and ApeKI digestions revealed low potential coverages for *Brassica napus*.

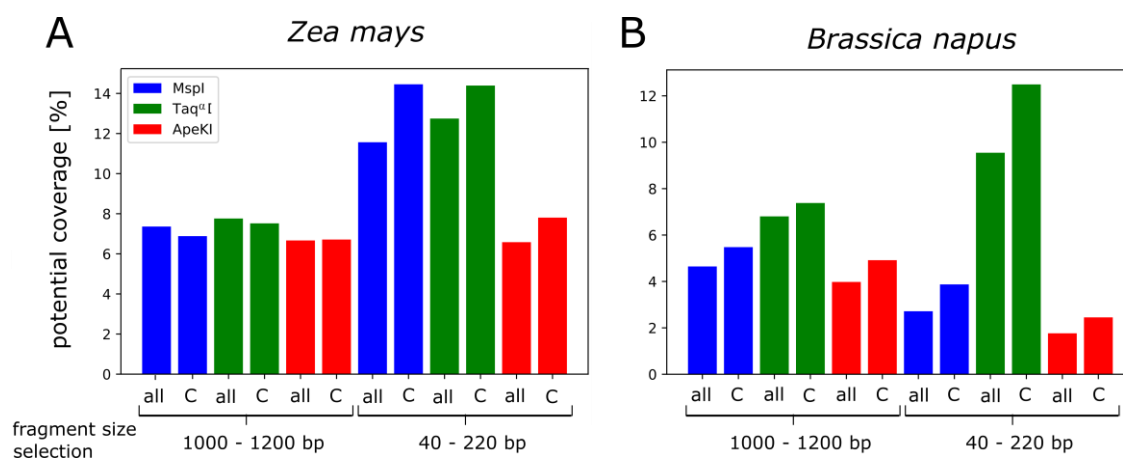


Figure 3.3: *In silico* restriction and fragment size selection of the used species.

The *Zea mays* (A) and *Brassica napus* (B) genomes were digested using the restriction enzymes MspI (blue), Taq^I (green) and ApeKI (red) used for RRBS tests with a size selection of 1000-1200 bp and 40-220 bp potentially generated through fragmentation by bisulfite conversion. Neighboring, similarly colored bars represent the relative coverage of the genome of all bases (all) and of only cytosines (C).

3.1.2 Comparison of two RRBS protocols

To determine the optimal RRBS library preparation method to examine a large genome and multiple samples of a population two RRBS protocols were compared: 1. The NextFlex Bisulfite Sequencing Kit (Bioo Scientific Corporation, Austin, USA) was tested following the corresponding manual. 2. A BS-seq protocol developed by Smallwood *et al.* (2014) was modified and adapted to RRBS (3.1.1, Edelmann and Scholten, 2018) which is further referred to as “Smallwood method”. Sequencing of two *Brassica napus* genotypes (4, 5) with four replicates for the Smallwood method and three replicates for the Nextflex method was performed for each protocol to compare the results of the different methods with the same genotypes (2.2.2.1. b). Sequencing data analysis of the sequencing was performed by Dr. Felix Seifert. For both strategies a small number of reads were lost due to adapter contamination (Table 3.1) and all reads exhibited a similarly high quality. Differences were observed in read duplication rate (Figure 3.4 A) and GC content resulting in a differential overall methylation rate in the samples from the same genotype (Table 3.1).

Table 3.1: Summary of the sequencing results before and after adapter trimming.

The methylation rate mean of the different cytosine contexts is given in the column of the respective context.

	sample	Number of raw sequence reads	Number of trimmed sequence reads	GC content [%]	CpG methylation rate [%]	CHG methylation rate [%]	CHH methylation rate [%]
Smallwood protocol	4a	22,446,450	21,179,977	26	48.7	21.0	11.5
	4b	19,143,567	180,600,085	25	43.8	16.6	8.6
	4d	16,834,846	15,918,080	25	46.6	20.3	9.9
	4e	21,741,001	20,513,654	25	55.0	23.9	12.7
	5a	19,639,779	18,545,665	26	46.5	19.3	10.8
	5b	20,891,155	19,706,663	24	45.2	18.2	9.2
	5c	18,491,184	17,477,122	26	52.2	21.4	11.9
	5d	18,105,225	17,088,662	25	50.5	19.8	9.7
NextFlex Kit	4-1	17,331,751	17,331,333	36	74.8	33.7	9.6
	4-2	18,539,002	18,538,584	37	82.2	41.0	12.1
	4-3	23,603,624	23,603,029	32	60.1	24.9	5.9
	5-1	17,561,426	17,560,966	34	63.4	28.1	7.7
	5-2	16,734,501	16,734,093	35	71.9	34.7	10.2
	5-3	19,369,818	19,369,381	37	81.8	44.0	14.7

The Smallwood method revealed less read duplication compared to the NextFlex method (**Figure 3.4 A and B**). The per base sequence content was more evenly distributed in the samples prepared with the Smallwood method compared to the NextFlex Kit (**Figure 3.4 C and D**). Both of the preparation methods revealed a base bias to a certain sequence (in Smallwood libraries the first (5'-) nine bases due to the random primers and in Nextflex libraries the first (5'-) three bases due to leftover of restriction enzyme sites), which were removed by trimming the appropriate number of bases of the sequencing reads. The libraries prepared with the NextFlex protocol showed higher overall methylation rate of especially CG and CHG contexts (CG: 60-82 %, CHG: 25-44 %, CHH: 6-15 %) compared with the Smallwood libraries (CG: 44-55 %, CHG: 17-24 %, CHH: 9-13 %, **Table 3.1**).

Sequence duplication levels

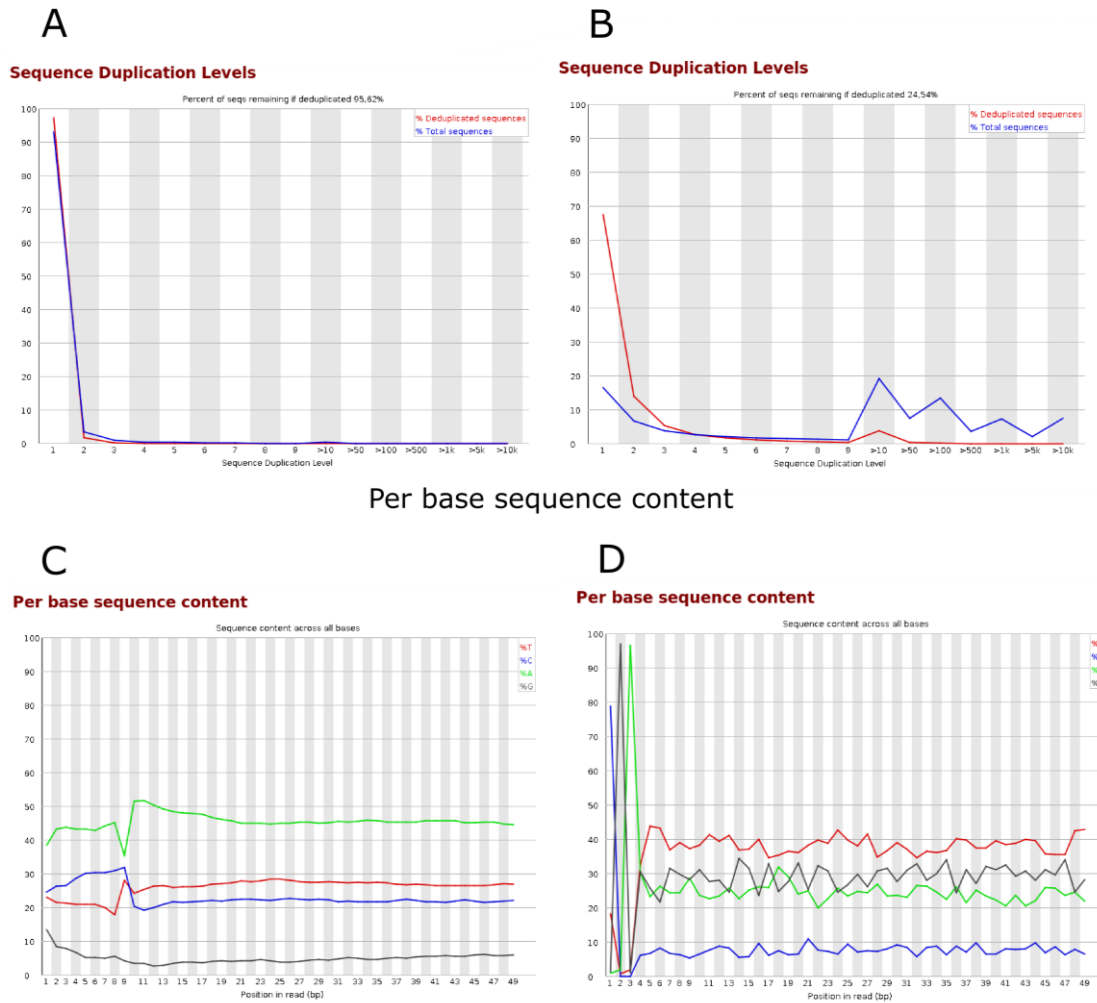


Figure 3.4: FastQC (v0.11.4) analysis of RRBS sequence reads (performed by Dr. Felix Seifert).

Read duplication (**A**) and Per base sequence content (**B**) are shown exemplary for the Smallwood (sample 4a) and the NextFlex (sample 4-1) protocol.

3.2 DNA methylation impact on heterosis during early embryogenesis in maize

DNA methylation has been shown to be involved in heterosis formation. Maize inbred and hybrid embryos were treated with the methyltransferase inhibitor 5-aza-2'-deoxycytidine (aza) to investigate epigenetic differences and their timing that manifest in hybrids. Growth rates of the resulting inbred and hybrid seedlings were measured and heterosis was calculated.

3.2.1 Confirmation of successful DNA demethylation through aza-treatment

To confirm the successful demethylation of the embryonic maize DNA in the tissue culture, RRBS was performed and sequencing results were analyzed. **Table 3.2** summarizes the results of the sequencing.

Table 3.2: Summary of the sequencing read quality and quantity of maize samples.

Methylation means and cytosine coverage were calculated with 10 reads/cytosine.

sample	number of trimmed reads	conversion rate (% cp)	mappability efficiency of unique reads (%)	CpG methylation rate [%]	CHG methylation rate [%]	CHH methylation rate [%]	cytosine coverage [%]
aza-treated embryos	51,136,525	99.2	16.1	45.23	43.63	1.54	0.76
non-treated embryos	25,593,611	99.1	15.2	58.46	56.98	1.82	0.09
aza-treated seedling	5,748,399	99.3	23.8	50.28	41.84	1.32	1.4
non-treated seedling	18,534,804	99.3	26.3	58.43	47.18	1.35	0.09

The RRBS sequencing yielded in 18-30 million 100 bp paired-end reads. Among those, 8.2 million reads (16.1%, aza-treated embryos) and 3.9 million reads (15.2%, non-treated embryos) as well as 12.3 million reads (23.8%, seedlings from aza-treated embryos) and 4.9 million reads (26.3%, seedlings from non-treated embryos) aligned to unique locations of the B73 reference genome (2.2.3.1). Treated and non-treated samples yielded in different read numbers, since for treated samples, sequencing reads of two replicates were pooled, whereas non-treated reads resulted from one library each. The mapping efficiencies differed between the samples and were overall low as a result of a genetically distant reference genotype the reads were mapped to as well as the bisulfite conversion.

Methylated cytosines were called for cytosines with a coverage of at least 10 reads which yielded in cytosine coverages of total cytosines from 0.1 to 1.4 %. Analysis of the cytosines in the different contexts revealed cytosine methylation frequencies of 45.23 % for CpG, 43.63 % for CHG, and 1.54 % for CHH contexts in the aza-treated embryos compared to 58.46 % for CpG, 56.98 % for CHG, and 1.82 % for CHH in the non-treated embryos.

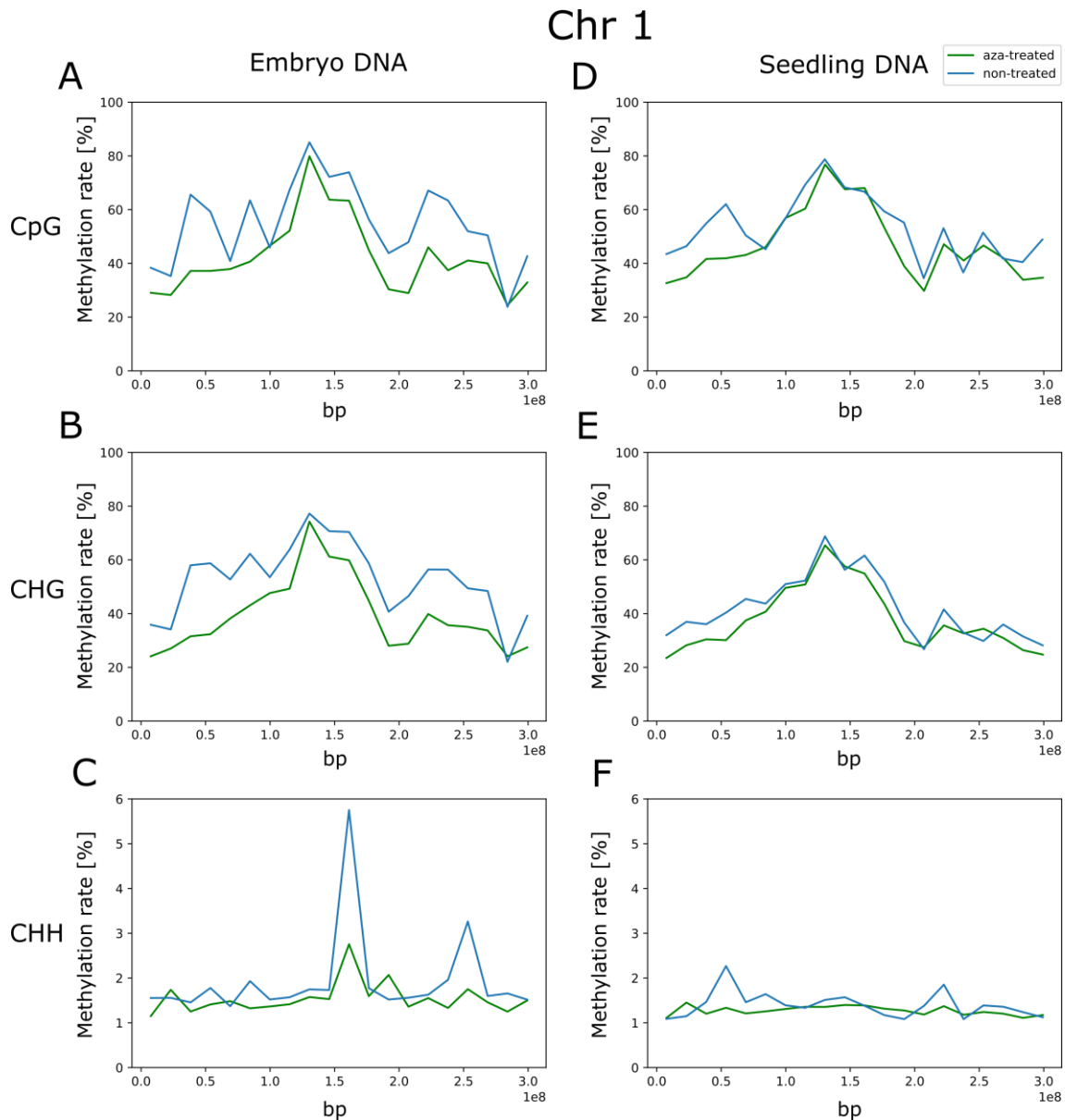


Figure 3.5: Methylation distribution patterns of aza-treated and non-treated embryonic DNA.

DNA methylation distribution is shown exemplary on the first maize chromosome of **A.** CpG contexts, **B.** CHG and **C.** CHH contexts. Methylation distribution patterns of aza-treated and non-treated seedling

DNA exemplars on the first chromosome are displayed in **A.** for CpG contexts, **B.** for CHG and **C.** for CHH contexts.

In summary, the data revealed significantly less DNA methylation in the aza-treated embryos in every cytosine context (CpG, CHG and CHH, $p < 0.05$, Student's *t* test), which is shown exemplarily on the first maize chromosome in **Figure 3.5 A, B and C.** The DNA of the seedlings grown from these embryos was significantly hypomethylated for CpG (50.28 % with aza-treatment, 58.43 % in non-treated, $p < 0.05$) and CHG (41.84 % with aza-treatment, 47.18 % in non-treated, $p < 0.05$) contexts through the aza-treatment, which is shown exemplarily on the first maize chromosome in **Figure 3.5 D, E and F.**

3.2.2 Effects of the DNA demethylation during embryo development on seedling growth rate

After validating that aza affects the methylation of DNA in early embryos, testing whether global demethylation influences heterosis formation in post-germination plants was performed. The growth rates of seedlings during their phase of linear growth after germination were determined for plants from aza-treated and non-treated embryos. Comparison of inbred lines and hybrids revealed the effect of embryonic demethylation on heterosis (**Figure 3.6**). As A188 represents the parent with superior growth performance *in vitro* (**Figure 3.6 A**), calculations of heterosis for growth rate performance refer to better-parent heterosis (BPH, **Figure 3.6 B**).

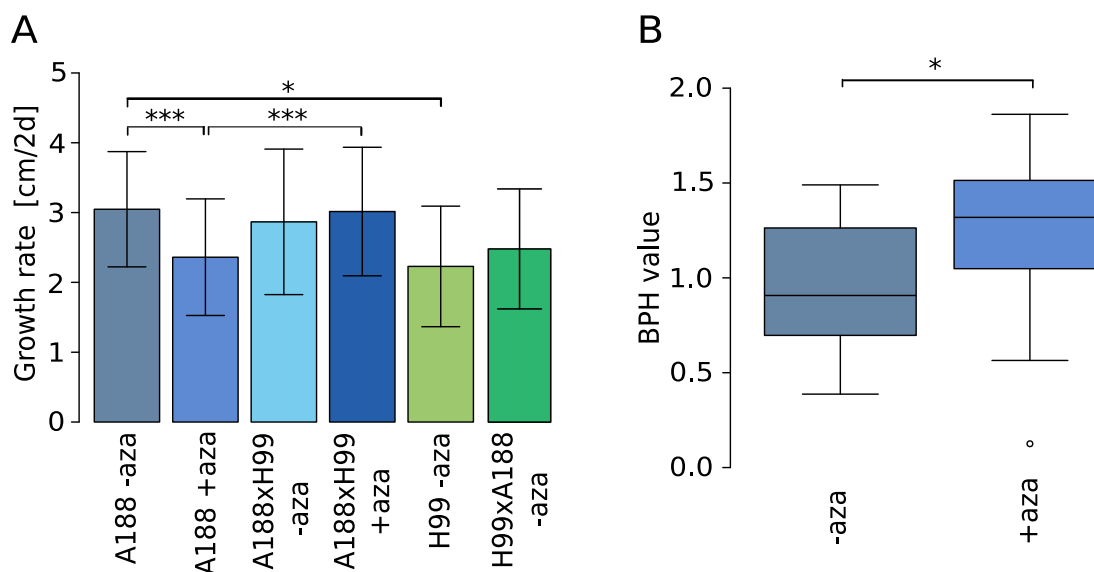


Figure 3.6: Growth rates and heterosis calculations of maize inbred lines and hybrids grown from aza-treated and non-treated embryos (1 dap).

A. Growth rate of inbred lines and F1 hybrids (A188 x H99), derived from seedlings grown *in vitro* from 1 day old aza-treated (+aza) and non-treated (-aza) embryos. Number of seedlings measured: 67 for non-treated A188 inbred lines, 10 for non-treated H99 inbred lines, 26 for non-treated A188 x H99 hybrids, 9 for non-treated H99 x A188 hybrids, 50 for aza-treated A188 inbred lines, 24 for aza-treated A188 x H99 hybrids. **B.** Growth rate better parent heterosis (BPH) calculation for aza-treated and non-treated F1 hybrids (A188 x H99) grown from *in vitro* culture of 1 dap embryos. * indicates p-value<0.05, *** indicates p-value<0.01 for pairs of samples (Student's t test).

Non-treated hybrids (A188 x H99, H99 x A188) showed no BPH for growth rate (**Figure 3.6 A, Table A.4**). The mid-parent heterosis (MPH) value was also relatively low (8.6%). Growth rates of aza-treated A188 inbred lines were significantly lower ($p<0.01$) compared to non-treated inbred lines, whereas aza-treated hybrids revealed no significant difference in growth rate compared to non-treated hybrids (**Figure 3.6 A**). The difference of growth rates between aza-treated inbred lines and hybrids was significant ($p<0.01$) and BPH of growth rate revealed a value for aza-treated lines of 26%. Comparing the BPH values of aza-treated and non-treated plants showed a significant increase of heterosis for growth rate of seedlings ($p<0.01$, **Figure 3.6 B, Table A.4**).

3.3 DNA methylation patterns in an oilseed rape population

To further characterize the impact of DNA methylation on the heterosis phenomenon, a population with paternal and maternal lines and their hybrids was analyzed (2.1.1). Therefore phenotypic (2.1.5.1) and methylation data (2.2.2) of a *Brassica napus* population subset consisting of 65 lines (**Table 2.1**) was collected.

3.3.1 Phenotypic variation in the *Brassica napus* population

To examine whether DNA methylation differences are correlating with phenotypic traits of agronomical relevance directly or whether a relation between DNA methylation in parents or hybrids and hybrid performance or heterosis, respectively, can be observed, field traits taken by collaborating breeders were included in the analyses. **Figure 3.7** shows all field trait performances of the parental inbred lines and the hybrids (**Table A.1, Table A.2**). A highly significantly improved hybrid performance was observed for field emergence, plant length, yield and glucosinolate content (**Table 3.3, Figure 3.7 A, D, F and H**), The latter should be as low as possible due to digestion problems of animals when used as feed. For oil content and thousand kernel weight (**Figure 3.7 E, G**), the paternal parent performed slightly better, while for beginning of flowering and protein contents in seeds the maternal parents showed higher performance (**Figure 3.7 B, C, Table 3.3**). The parents did not significantly differ in any trait. In most traits, the hybrids' performance was significantly aberrant from both, maternal and paternal lines in one direction or the other.

Table 3.3: Averages and standard deviations of the analyzed phenotypic *in vitro* seedling root/shoot as well as field traits per maternal lines, paternal lines and hybrids.

trait	hybrids		maternal lines		paternal lines	
	Mean	STD	Mean	STD	Mean	STD
Total root length (cm)	176.88	47.36	98.32	25.72	96.22	31.56
Total root growth rate (cm/4d)	12.44	3.40	6.91	1.82	6.75	2.25
Taproot length (cm)	31.79	5.93	24.96	4.78	25.05	3.99
Taproot growth rate (cm/4d)	2.19	0.42	1.73	0.32	1.76	0.28
Lateral root length (cm)	144.08	41.52	73.19	21.67	71.03	28.74
Lateral root growth rate (cm/4d)	10.19	2.99	5.17	1.55	4.99	2.06
Root system depth (cm)	28.15	4.07	23.02	4.38	23.21	3.45
Root system width (cm)	10.43	2.09	8.23	1.40	7.06	1.68
Convex hull area (cm²)	179.16	58.74	107.81	34.15	95.60	35.13
Lateral root branching angle (cm)	71.60	5.07	72.48	6.10	69.70	8.33
Number of laterals	91.18	25.41	57.19	17.55	58.63	14.32
Shoot dry weight (mg)	11.51	2.52	6.40	1.53	6.15	1.29
Root dry weight (mg)	4.59	0.99	2.47	0.56	2.87	0.96
Root-to-shoot ratio	0.40	0.06	0.40	0.10	0.48	0.17
Thousand kernel weight (mg)	4.79	0.23	5.16	0.45	4.80	0.23
Yield (dt/ha)	41.30	2.14	38.57	4.89	36.69	2.22
Plant length (cm)	167.55	3.50	159.83	6.12	159.82	5.54
Beginning of flowering (d)	110.23	1.18	111.85	2.01	112.81	2.26
Field emergence (d)	7.80	0.29	7.35	0.45	7.43	0.26
Oil content (%)	49.59	0.88	49.33	1.25	48.83	1.29
Glucosinolate content (%)	12.39	1.04	13.10	1.94	14.10	2.09
Protein content in seeds (%)	19.32	0.31	20.12	0.69	20.37	0.35

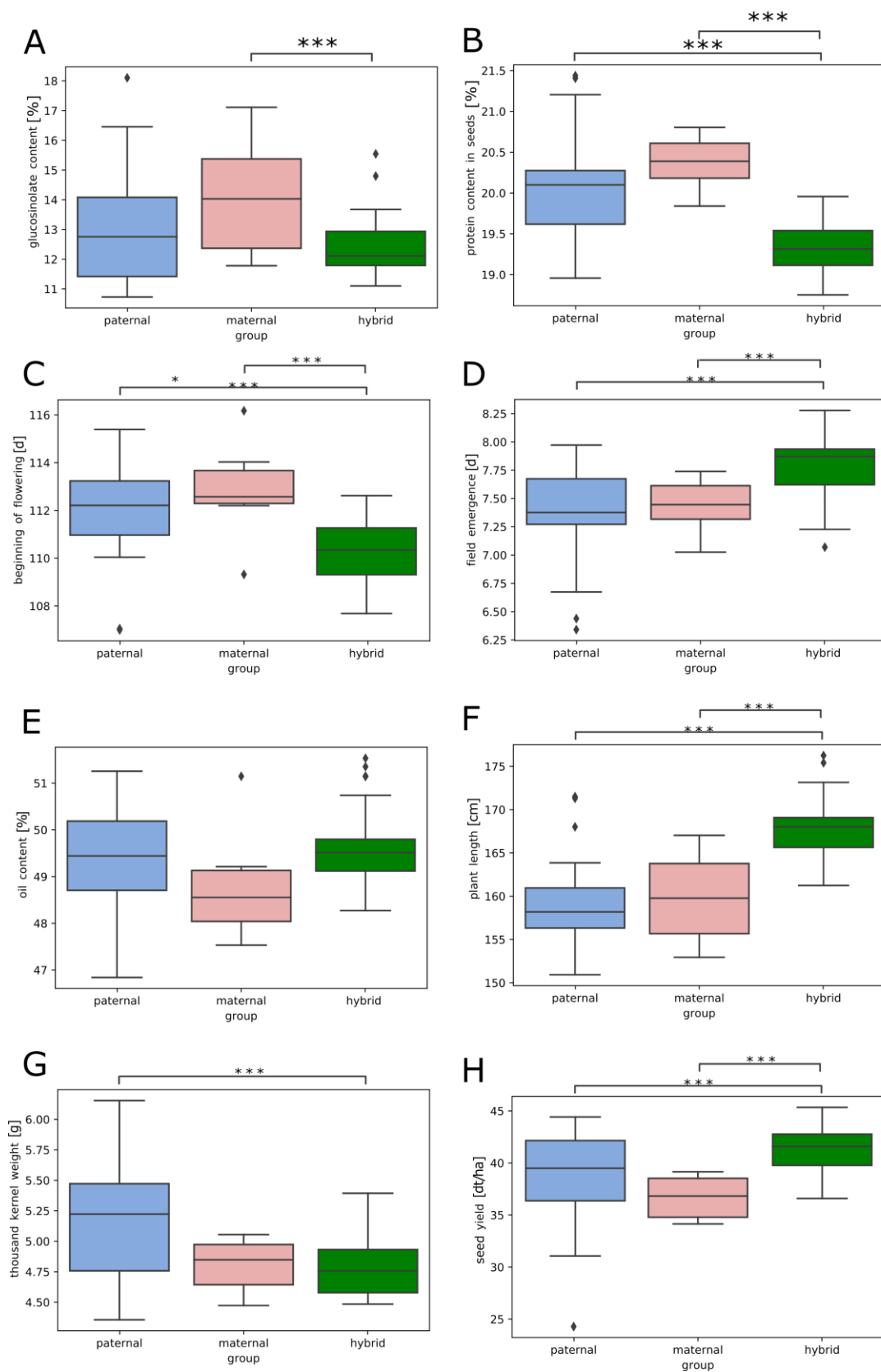


Figure 3.7: Phenotypic variability of field traits.

A. glucosinolate content, **B.** protein content in seeds, **C.** beginning of flowering, **D.** field emergence, **E.** oil content, **F.** plant length, **G.** thousand kernel weight and **H.** seed yield separately for paternal lines (n=21), maternal lines (n=6) and hybrids (n=27, *** p-value<0.01, Student's t test).

To analyze the DNA methylation influence on heterosis of traits measured at the same stage as molecular data, root and shoot traits measured in *in vitro* seedlings by a collaborating institute were used. In this stage and data, hybrid performance was higher compared to the inbred parents as in the field data (**Figure 3.7**, **Figure 3.8** and **Figure 3.9**).

Hybrid performance significantly increased for total root length, taproot length and lateral root length (**Figure 3.8 A, D and F**) and as well as the resulting growth rates (**Figure 3.8 B, E, and G, Table 3.3**). MPH was calculated for the representative trait total root length growth rate (**Figure 3.8 C**), showing a MPH value of 0.6 – 2.7.

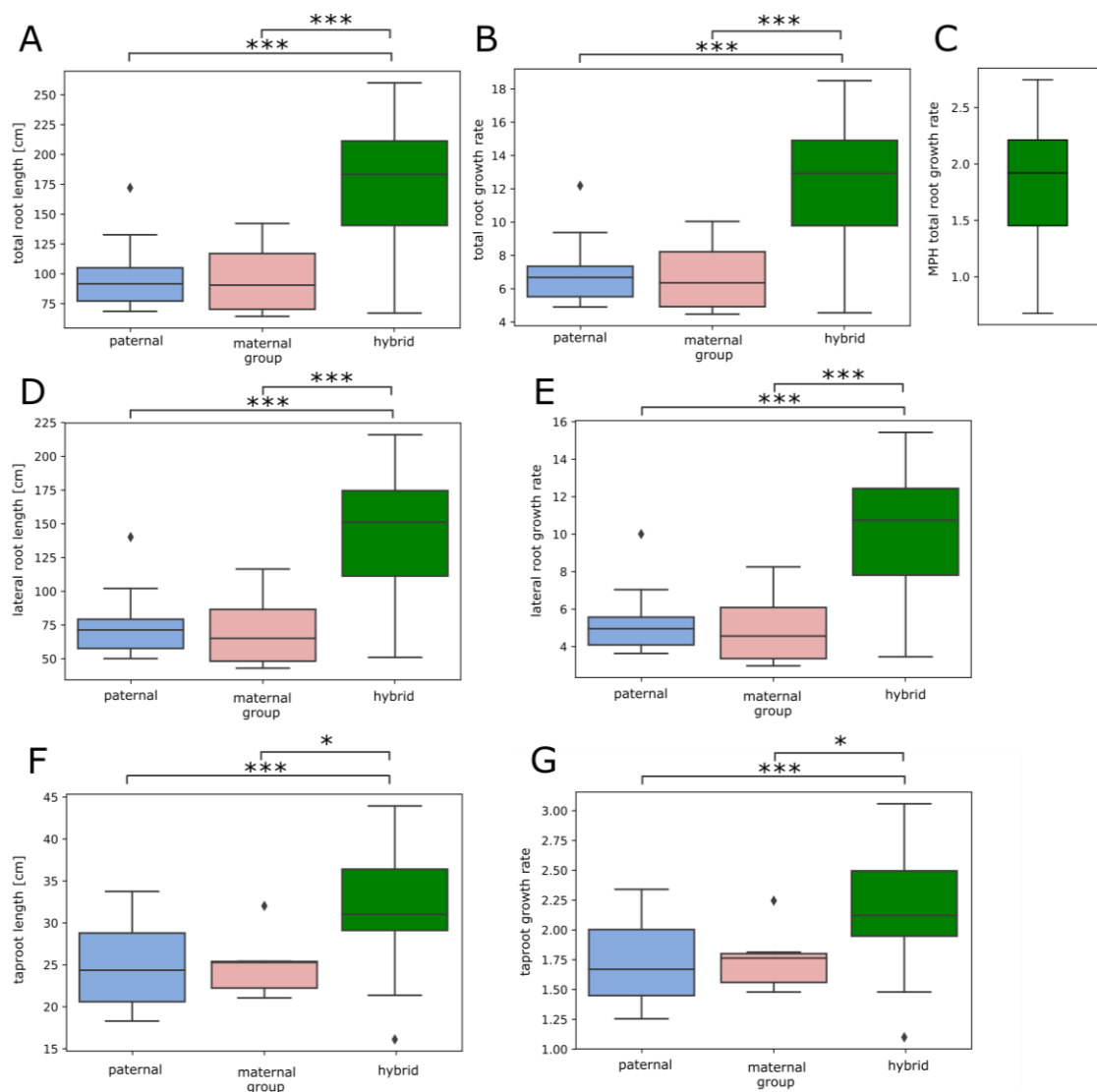


Figure 3.8: Phenotypic variability of *in vitro* seedling root lengths.

A. total root length, **B.** total root growth rate, **C.** MPH of total root growth rate, **D.** lateral root length, **E.** lateral root growth rate, **F.** taproot length and **G.** taproot growth rate separately for paternal lines (n=21), maternal lines (n=6) and hybrids (n=27, *** p-value<0.01, * p-value<0.05, Student's t test).

Root system depth (RSD, **Figure 3.9 A**), root system width (RSW, **Figure 3.9 B**), convex hull area (CHA, **Figure 3.9 C**), number of laterals (NRLR, **Figure 3.9 E**), shoot dry weight (SDW, **Figure 3.9 F**) and root dry weight (RDW, **Figure 3.9 G**) showed a significant performance increase in hybrids, too (**Figure 3.9**), whereas lateral root branching angle (BA) and root-to-shoot ratio (RS) did not differ from their parents (**Figure 3.9 D and H**). Parental values were in a similar range (**Table 3.3, Figure 3.9**).

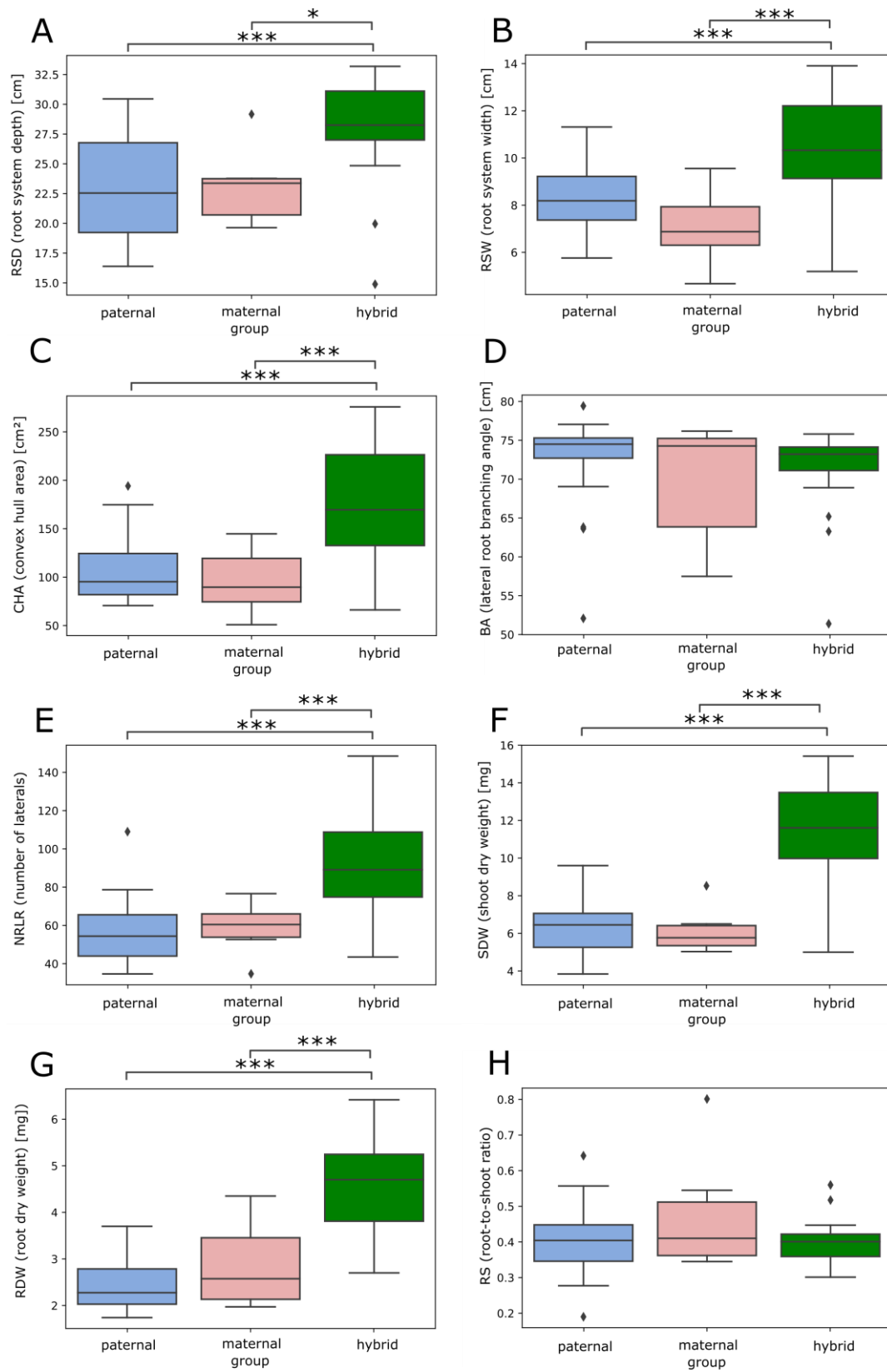


Figure 3.9: Phenotypic variability of *in vitro* seedling root lengths.

A. root system depth, **B.** root system width, **C.** convex hull area, **D.** lateral root branching angle, **E.** number of laterals, **F.** shoot dry weight, **G.** root dry weight and **H.** root-to-shoot ratio separately for paternal lines (n=21), maternal lines (n=6) and hybrids (n=27, *** p-value<0.01, * p-value<0.05, Student's t test).

3.3.2 Summary statistics of DNA methylation

Table 3.4 summarizes the RRBS sequencing results as means of the maternal lines, paternal lines and hybrids. Values for each genotype are listed in the appendix (**Table A.5**).

Table 3.4: Summary of the sequencing read quantity as well as methylation statistics for sequenced samples of rapeseed maternal lines, paternal lines and hybrids.

Methylation rates and cytosine coverages were calculated for all cytosines with at least 5 reads. CR = conversion rate calculation by cpDNA, ME = mapping efficiency, CpG = CpG methylation rate, CHG = CHG methylation rate, CHH = CHH methylation rate and CC = Cytosine coverage.

group	reads	CR [%]	ME [%]	CpG [%]	CHG [%]	CHH [%]	CC [%]
paternal lines	10,638,410 ± 1,916,198	99.23 ± 0.24	38.40 ± 4.32	46.21 ± 6.76	17.69 ± 2.66	4.53 ± 0.69	3.62 ± 1.69
maternal lines	10,194,682 ± 2,507,319	99.33 ± 0.09	37.07 ± 4.01	46.47 ± 8.28	17.99 ± 4.04	4.52 ± 1.05	3.19 ± 1.94
hybrids	10,481,730 ± 1,773,778	99.34 ± 0.11	36.73 ± 3.98	44.88 ± 6.30	17.05 ± 2.49	4.22 ± 0.52	3.14 ± 1.32

Sequencing of RRBS libraries yielded in about 6 - 16 million 100 bp paired-end reads. The mapping of these sequencing reads resulted in 31.3 – 44.3 % (maternal lines), 28.1 – 45.5 % (paternal lines) and about 26.5 – 44.4 % (hybrids) of total reads aligning to unique locations of the *Brassica napus* cultivar Drakkar reference genome (2.2.3.1). 0.8 to 8 % of all cytosines exhibited a coverage of at least 5 sequence reads. Analysis of the cytosines in the different contexts revealed cytosine methylation rates of the maternal lines ranging from 36.1 – 60.9 % for CpG, 14.1 - 25.6 % for CHG and 3.4 - 6.5 % for CHH. Paternal lines showed methylation rates of 38.6 – 63.8 % in CpG, 13.9 - 23.3 % in CHG and 3.4 - 6.2 % in CHH context. For hybrids, methylation rates of 38.1 – 65.2 % for CpG, 13.8 - 25.2 % for CHG and 3.2 - 5.7 % for CHH context were observed (**Table 3.4, Figure 3.10**).

A methylation rate mean difference of approximately 10 % between the subgenomes A and C were revealed at CG contexts (**Figure 3.10 A**), which was significant for paternal lines and hybrids (p<0.01, Student's t test). 3 % difference between A and C for

CHGs (**Figure 3.10 B**) and ca. 1 % difference for CHHs (**Figure 3.10 C**) were displayed, which were significant for paternal group and hybrids.

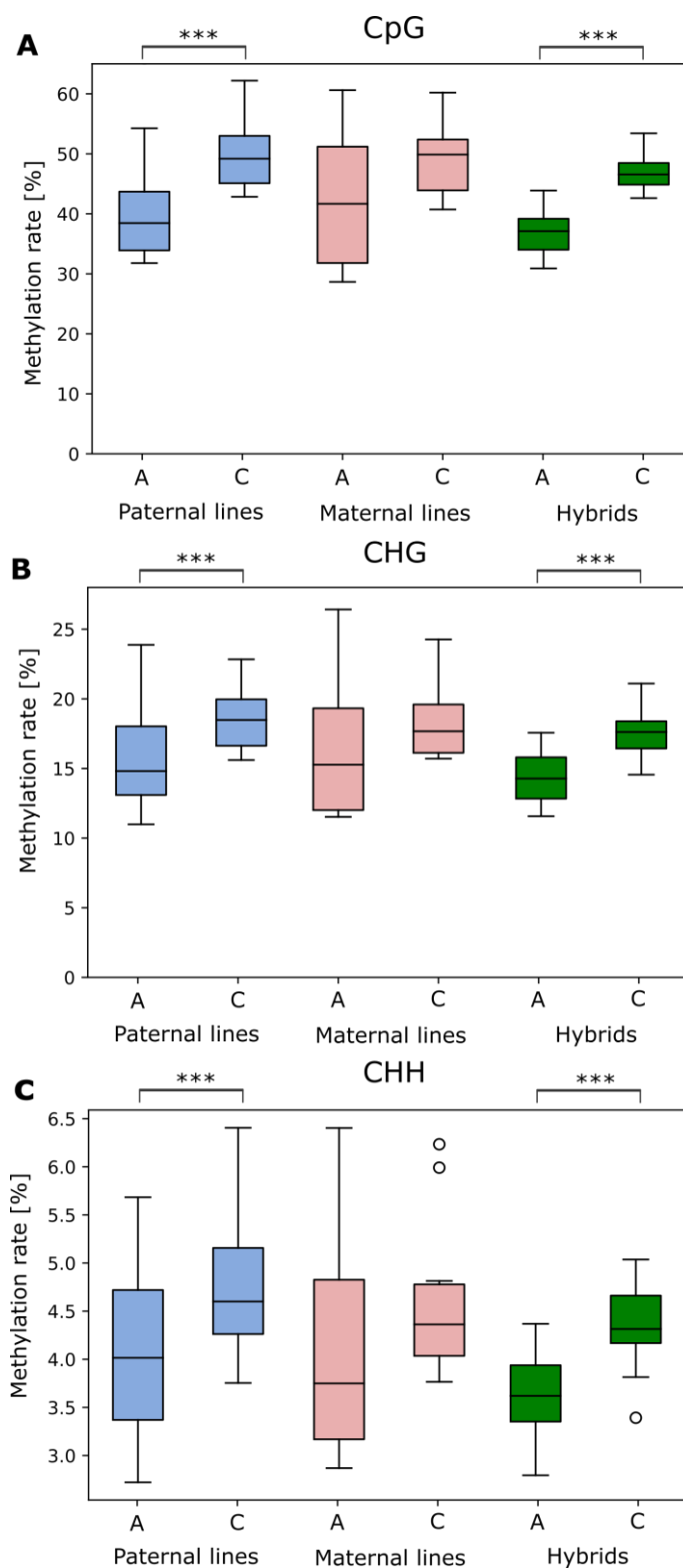


Figure 3.10: Methylation rates of paternal lines, maternal lines and hybrids. Methylation rates of **A.** CpG, **B.** CHG and **C.** CHH contexts are divided by subgenome A and C.

The average CpG, CHG, and CHH methylation rates were highly correlated (**Figure 3.11 A**). A moderate negative correlation between cytosine coverage and mean methylation per genotype (Pearson's $r < -0.6$, **Figure 3.11 A**) was observed, which would indicate that a higher coverage yielded in more sequence reads revealing low mean methylation (shown for CpG methylation in **Figure 3.11 B**). At a saturation of about 3 % cytosine coverage this effect was abolished. No strong correlation was observed between methylation rates and the count of trimmed sequencing reads and conversion rate, although the number of trimmed reads was strongly positively correlated with cytosine coverage (**Figure 3.11 A**). Mapping efficiency and cytosine coverages as well as methylation rates revealed weak correlations (**Figure 3.11 A**).

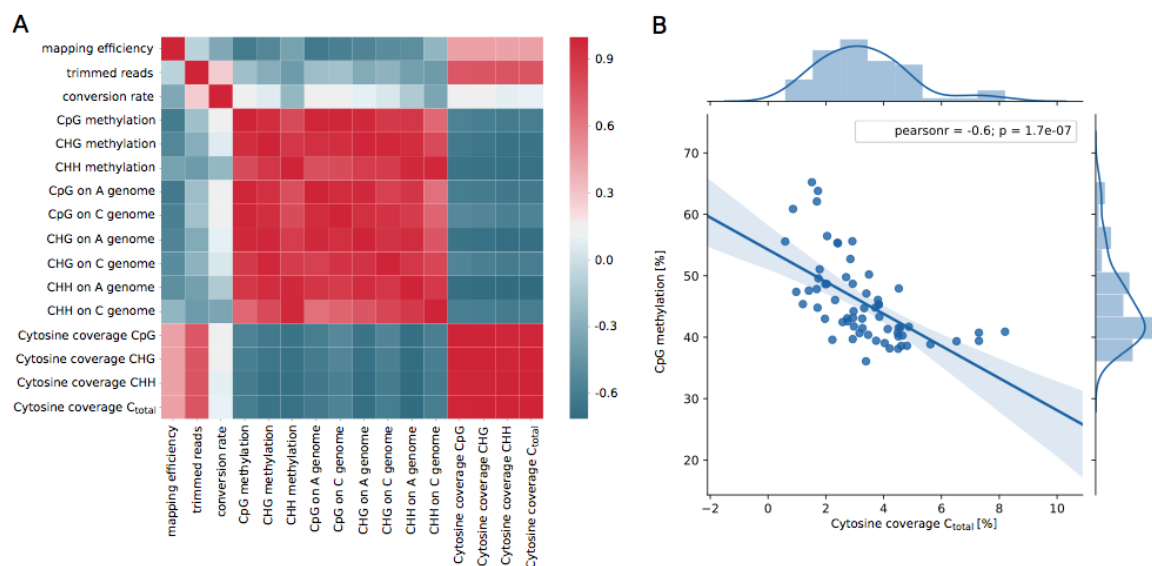


Figure 3.11: Correlation of cytosine coverage and mean DNA methylation rates.

A. Heatmap based on Pearson correlation coefficients between the methylation means of the contexts, cytosine coverage and mapping efficiency. **B.** Correlation of CpG methylation means and cytosine coverage.

As RRBS for a population requires a minimal coverage because of increased sample size, the methylation means apparently linked to coverage were included in the further analysis only in comparisons between groups, where this effect should be equal distributed. Regional analysis was further used to deplete the apparent influence of coverage on methylation.

3.3.3 Analysis of parental differential methylation

Differentially methylated positions (DMPs) and regions (DMRs) were called between the parents of each hybrid to analyze potential effects of parental epigenetic differences in hybridization and subsequently on heterosis.

Varying differences of methylation levels (ΔmC) were tested to define a differentially methylated region/position ranging from 10 % to 50 %. Several ΔmC were tested for DMRs and DMPs to find the most appropriate ratio between ΔmC of DMRs and DMR number (same for DMPs). The following figures show the result for all parental combinations where hybrid data were available (**Figure 3.12**, **Figure 3.13**).

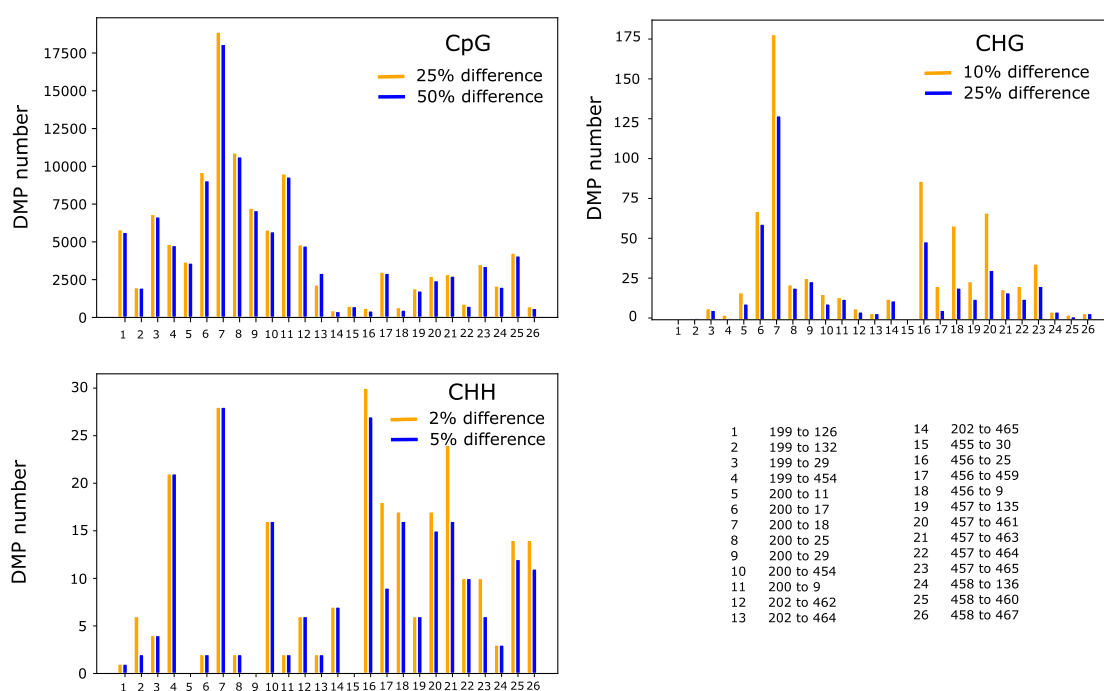


Figure 3.12: Number of differentially methylated positions per difference threshold for each cytosine context and hybrid line.

DMPs with a minimum differential methylation of 25 % and 50 % in CpG context, 10 % and 25 % in CHG context and 2 % and 5 % in CHH context. The numbers on the x axis indicate the parental combination.

The DMP numbers at 25 % ΔmCG were similar to 50 % ΔmCG in every sample ranging from 100 up to 17,700 DMPs (**Figure 3.12**). For the CHG and CHH context, the different ΔmCs , which were tested, were selected for the general methylation level of the respective context. As CHG contexts were methylated about 17 % (compared to CpG methylation with about 46 %, **Table 3.4**), $\Delta mCHGs$ of 10 % and 25 % were selected. DMP numbers in CHG contexts were comparatively low ranging from 0 in few samples

to 175 for 10 % in one parental comparison (**Figure 3.12**). Methylation levels in CHH contexts ranged mainly from 2-5 % (**Table 3.4**). Therefore, ΔmCHH s of 2 % and 5 % were selected for tests. ΔmCHH increase had little or no effect (**Figure 3.12**).

In contrast, DMR numbers significantly decreased ($p < 0.05$, Student's t test) with increasing differential methylation threshold (ΔmCG 25 % and 50 %, ΔmCHG 10 % and 25 % and ΔmCHH 2 %, 5 % and 25 %). For CG, the number of DMRs diminished about half from ~ 1000 to ~ 500 on average (**Figure 3.13**). Increasing the differential methylation threshold from 10 % to 25 % in CHG context excluded about 80 % of the DMRs. The number of DMRs was reduced from ~ 8000 to ~ 1800 on average (**Figure 3.13**). For CHH, a methylation difference threshold of 2 % and 5 %, respectively, led to DMR numbers of up to 40,000, which was strongly decreased for $\Delta\text{mCHH} = 25\%$ (**Figure 3.13**).

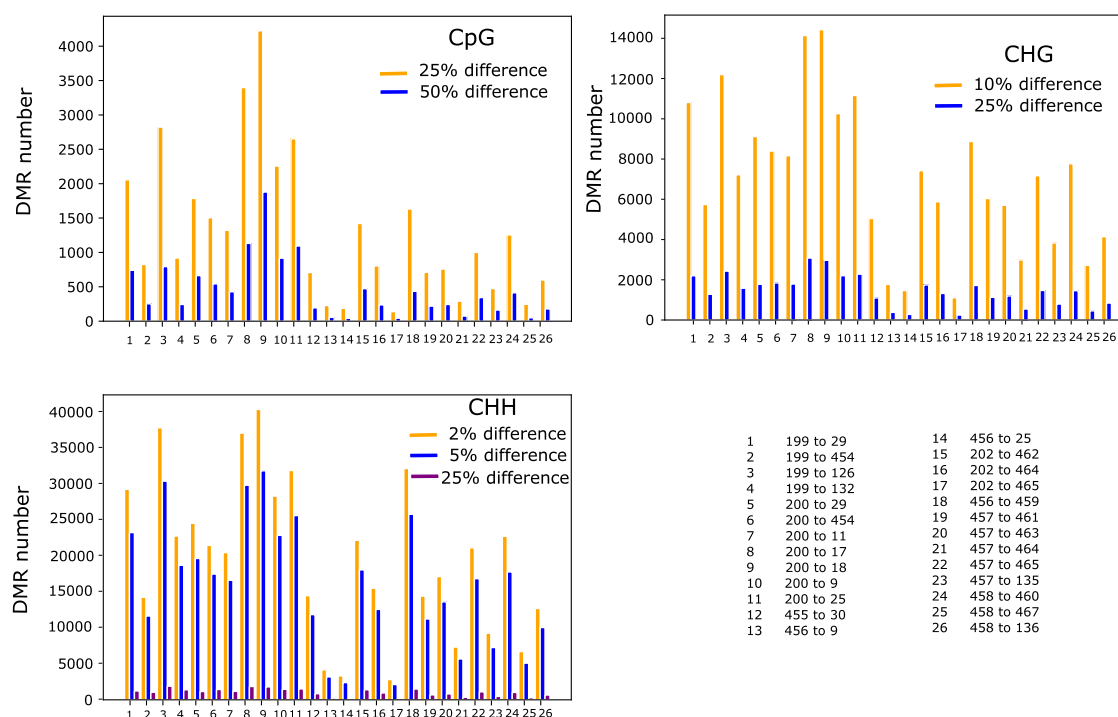


Figure 3.13: Number of differentially methylated regions per difference threshold for each cytosine context and hybrid line.

DMRs with a minimum differential methylation of 25 % and 50 % in CpG context, 10 % and 25 % in CHG context and 2 % and 5 % and 25 % in CHH context. The numbers on the x axis indicate the parental combination.

Since for CHG and CHH, the number of DMPs available for further analysis was very small for either ΔmCs (4.2.2.1), all analyses concerning differences between parents were performed with DMRs ($\Delta\text{mC} = 25\%$ for all contexts).

Pearson correlation analysis was conducted to examine the relations between DNA methylation, differential DNA methylation of regions and positions, and coverage. **Figure 3.14** shows the correlation coefficients: Whereas the cytosine coverage revealed a negative correlation with the genome wide DNA methylation levels (red arrow in **Figure 3.14**), no correlations were detected between coverage and all DMR and DMP features. Strong positive correlations were exhibited between the methylation characteristics of the CG context ($r > 0.9$), except the mean of a differentially methylated region or position. The methylation features of CG additionally positively correlated with the differentially methylated regions and positions of both other contexts ($r > 0.6$), while the correlation of CHG and CHH DMRs and DMPs was lower ($r = 0.4 - 0.6$). However, the mean methylations of every context strongly correlated with each other ($r > 0.7$).

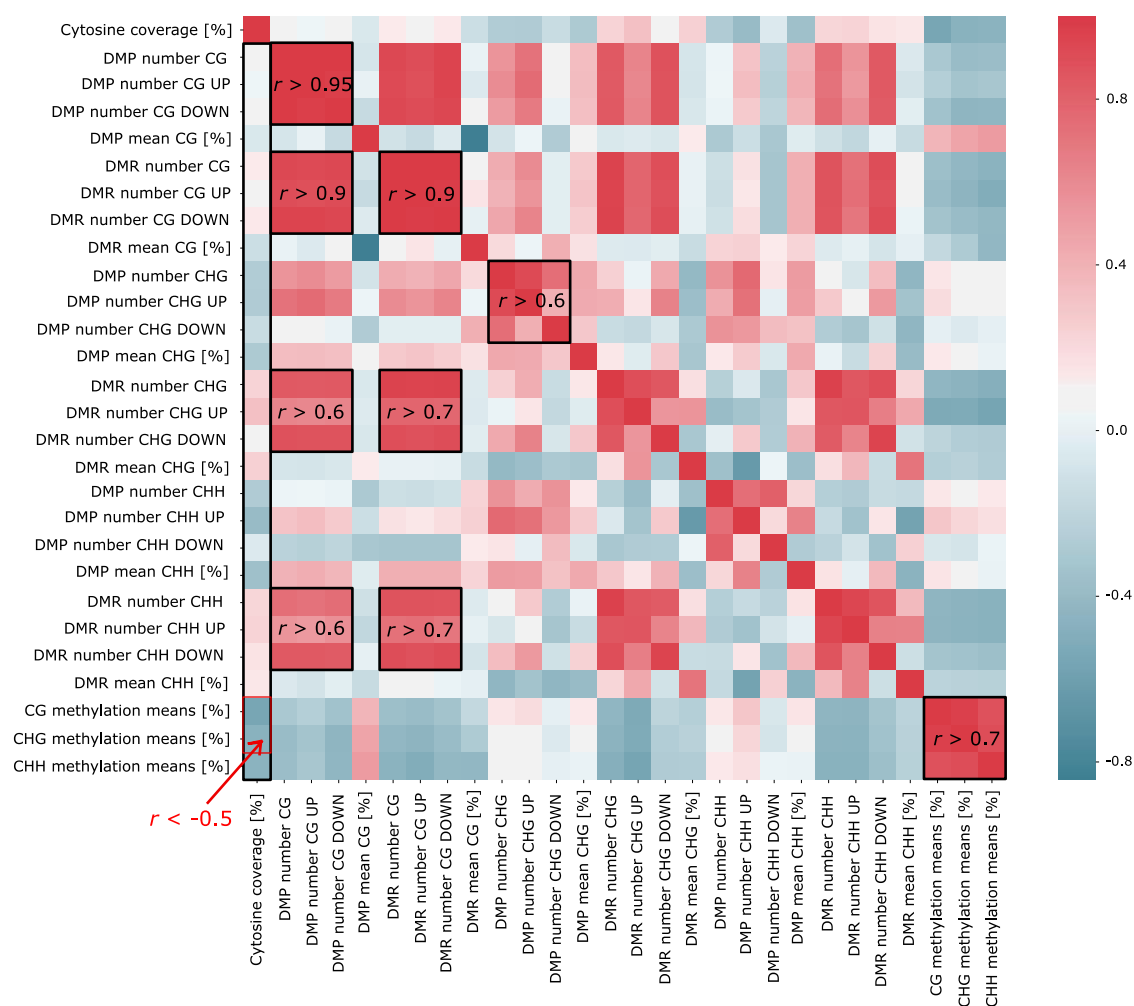


Figure 3.14: Heatmap based on Pearson coefficient shows correlations between means and numbers (up and down regulation and both) of the differentially methylated positions and regions of the three contexts.

All given r values are significant with $p < 0.01$. The red arrow indicates the negative correlation of methylation means and cytosine coverage.

The differentially methylated regions between parents were used to investigate the methylation inheritance patterns of maternal and paternal parents to their hybrids. The portion of additively, dominantly and overdominantly methylated regions was calculated in relation to the total number of parental DMRs additionally covered in hybrids. The contribution of the maternal and paternal parent to the hybrid methylation state was determined by comparing dominant and overdominant methylation levels in the hybrid with either maternal or paternal methylation level without consideration of either high or low methylation. In addition, trans-chromosomal methylation (TCM) and demethylation (TCdM) events were identified using the same approach but without consideration of maternal or paternal origin. Hybrid values were considered as additive within a deviation of 10 % methylation rate from the parental mean. Dominant patterns were defined with 10 % methylation rate deviation from one of the parental values. Overdominant was either below (for low parent) or above (for high parent) these values. The same approach was used for determination of trans-chromosomal methylation and demethylation events as well as maternal or paternal contribution.

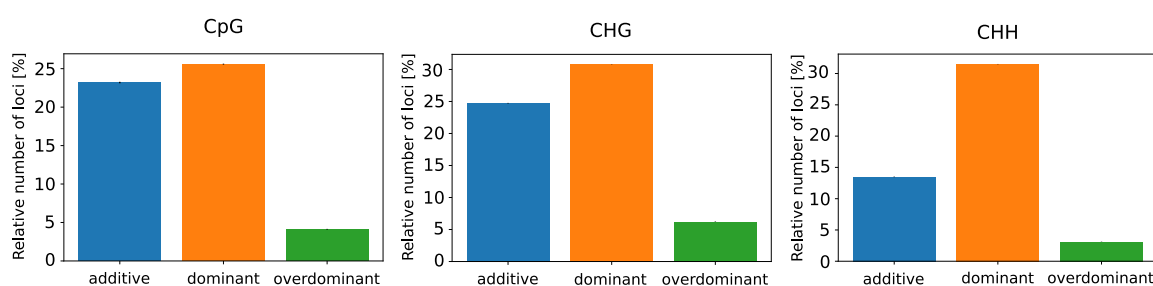


Figure 3.15: DNA methylation patterns in hybrids compared to parents in regions of parental differential methylation of all three cytosine contexts.

Black bars indicate standard deviations.

For all contexts, the majority of methylation patterns in the hybrids were dominant in parental DMRs (**Figure 3.15**) with a stronger maternal contribution for CHG and CHH contexts, while the paternal and maternal contributions are equal for CG contexts (**Figure 3.16**). Additive methylation in hybrids was approximately equal in CG and CHG contexts with ~23 %. In CHH, additive loci were relatively less compared to CG and CHG, while dominant methylation was similar in all contexts with about 30 % of the

loci. Overdominant methylation was rare compared to dominant and additive methylation patterns with less than 5 % of the included DMRs for every context (**Figure 3.15**).

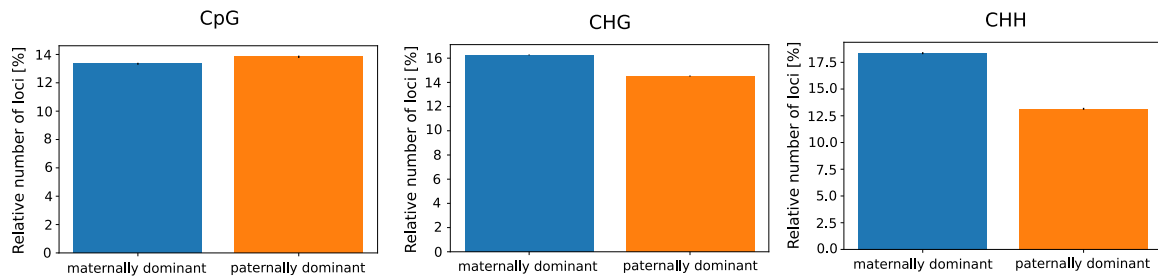


Figure 3.16: Adaption of the hybrid methylation level to either the maternal or paternal state in regions of parental differential methylation in hybrids of all three cytosine contexts.

Black bars indicate standard deviations.

TCM and TCdM events were examined to get an overview whether the hybrid is rather adapting to the higher or the lower methylation state of the parents. TCM and TCdM events were relatively equal in CG contexts, but found at a low number of CG DMRs (~15 %) compared to the other contexts (up to 31 % for CHH). For CHG and CHH, an increasing difference was observed between TCM and TCdM with a higher proportion of TCdM events (CHG: 22 %, CHH: 31 %) compared to TCM events (CHG: 15 %, CHH: 2.5 %, **Figure 3.17**).

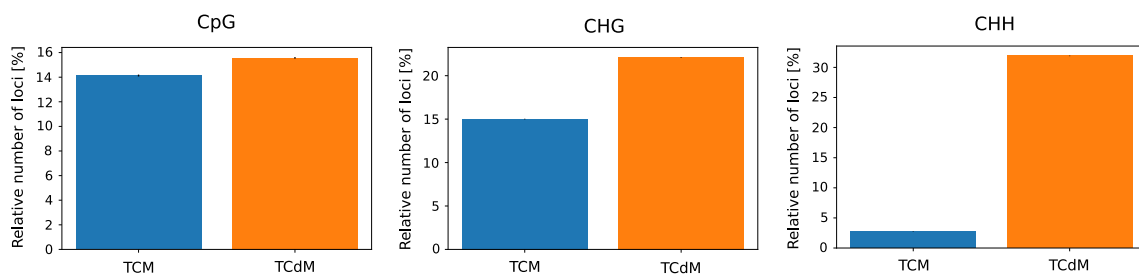


Figure 3.17: Trans-chromosomal DNA methylation (TCM) and demethylation (TCdM) events in regions of parental differential methylation in hybrids of all three cytosine contexts.

Black bars indicate standard deviations.

3.3.4 DNA methylation pattern analysis on genome annotations

Covered cytosines were mapped on annotated coding sequences (CDS) and transposable elements (TE, 2.2.3.3) to determine the importance of methylation differences between parents and hybrids on features to whose regulation methylation might contribute.

Most of the cytosines covered by RRBS sequencing were found neither in CDS nor TE sequences (60-75 %) but the remaining genome (**Figure 3.18**). For all contexts, an equal distribution between CDS and TEs was determined. Least coverage on annotated features was shown for CHG contexts, while CGs was increasingly distributed on TEs and CDS. CHHs were found the most of all contexts on annotated features.

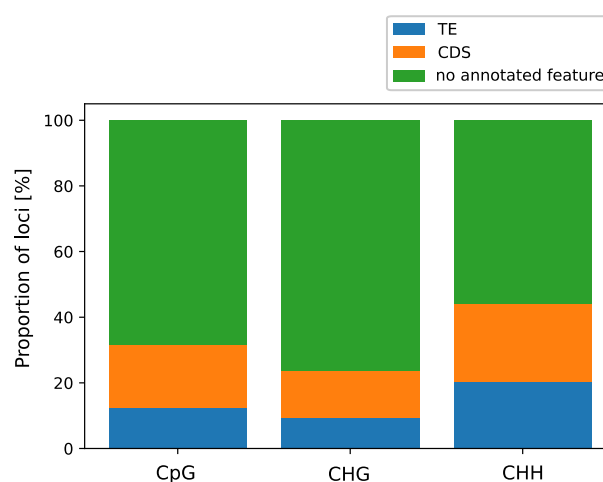


Figure 3.18: Cytosine distribution on CDS, TEs and remaining regions.

The cytosine methylation distribution (in bins including all covered loci of each sample) over CDS revealed a valley-like pattern (**Figure 3.19**). The gene body of all covered genes was less methylated in all contexts, groups and both the A and C genome compared to the flanking regions up- and downstream of the coding sequence (each 1 kb). The region within 1 kb upstream the CDS showed an increased methylation compared to 1 kb downstream. A significantly increased methylation of all cytosine contexts is exhibited between A and C genome of all groups (**Figure 3.19**), with average values rising about 8 % in CG, 2-4 % in CHG and 0.5 % in CHH contexts.

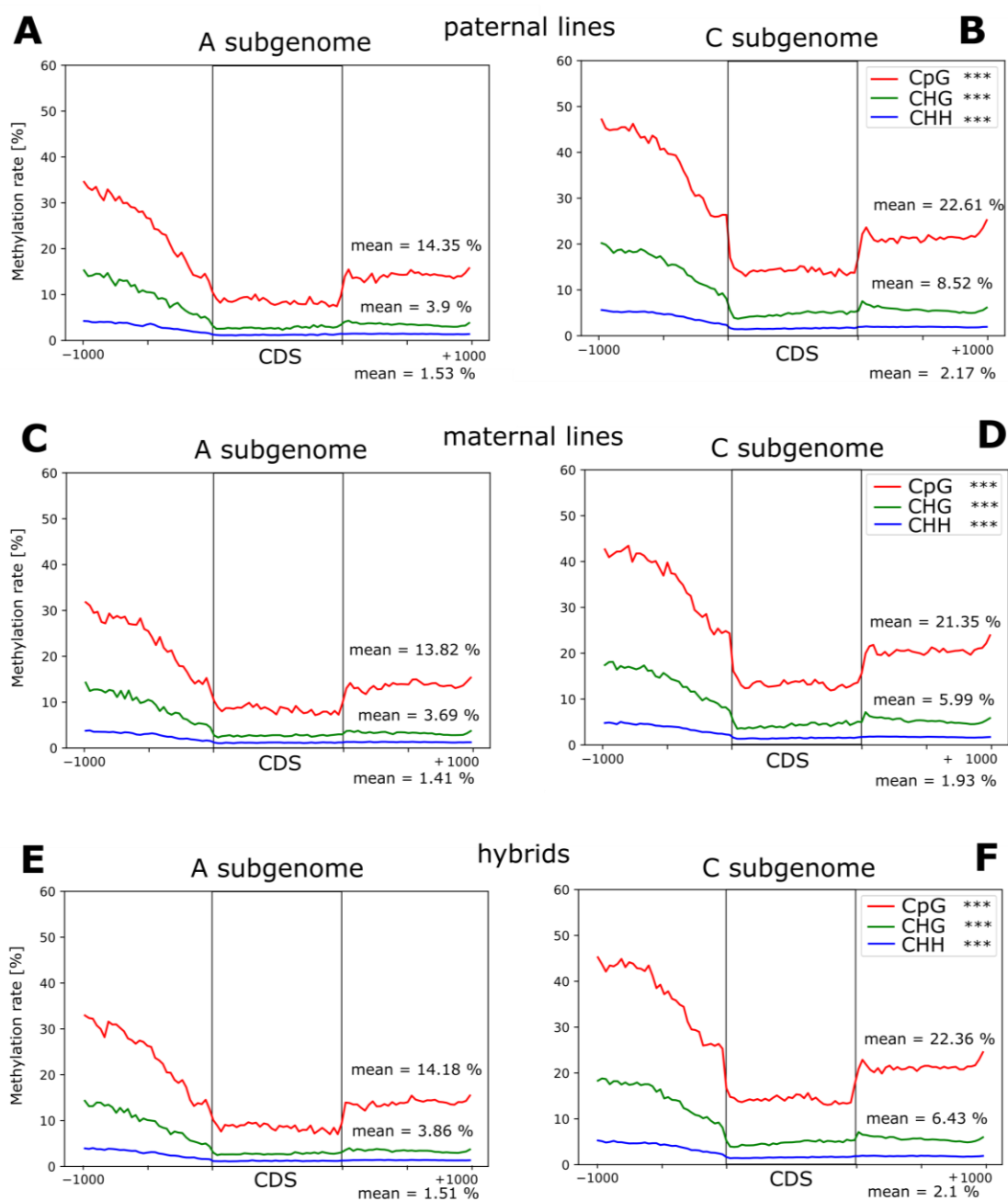


Figure 3.19: Cytosine methylation rate distribution in 100 bins over CDS including 1kb up- and downstream regions of *Brassica napus*.

Cytosine methylation distribution is shown for **A.** A subgenome and **B.** C subgenome of paternal lines, **C.** A subgenome and **D.** C subgenome of maternal lines and **E.** A subgenome and **F.** C subgenome of hybrid lines.

In contrast to CDS, cytosine methylation was strongly elevated in TE bodies (**Figure 3.20**). The significantly increased methylation of the C genome was observed in flanking

regions and transposable elements themselves. The region within 1 kb upstream the TEs showed, opposing to CDS, decreased methylation compared to 1 kb downstream.

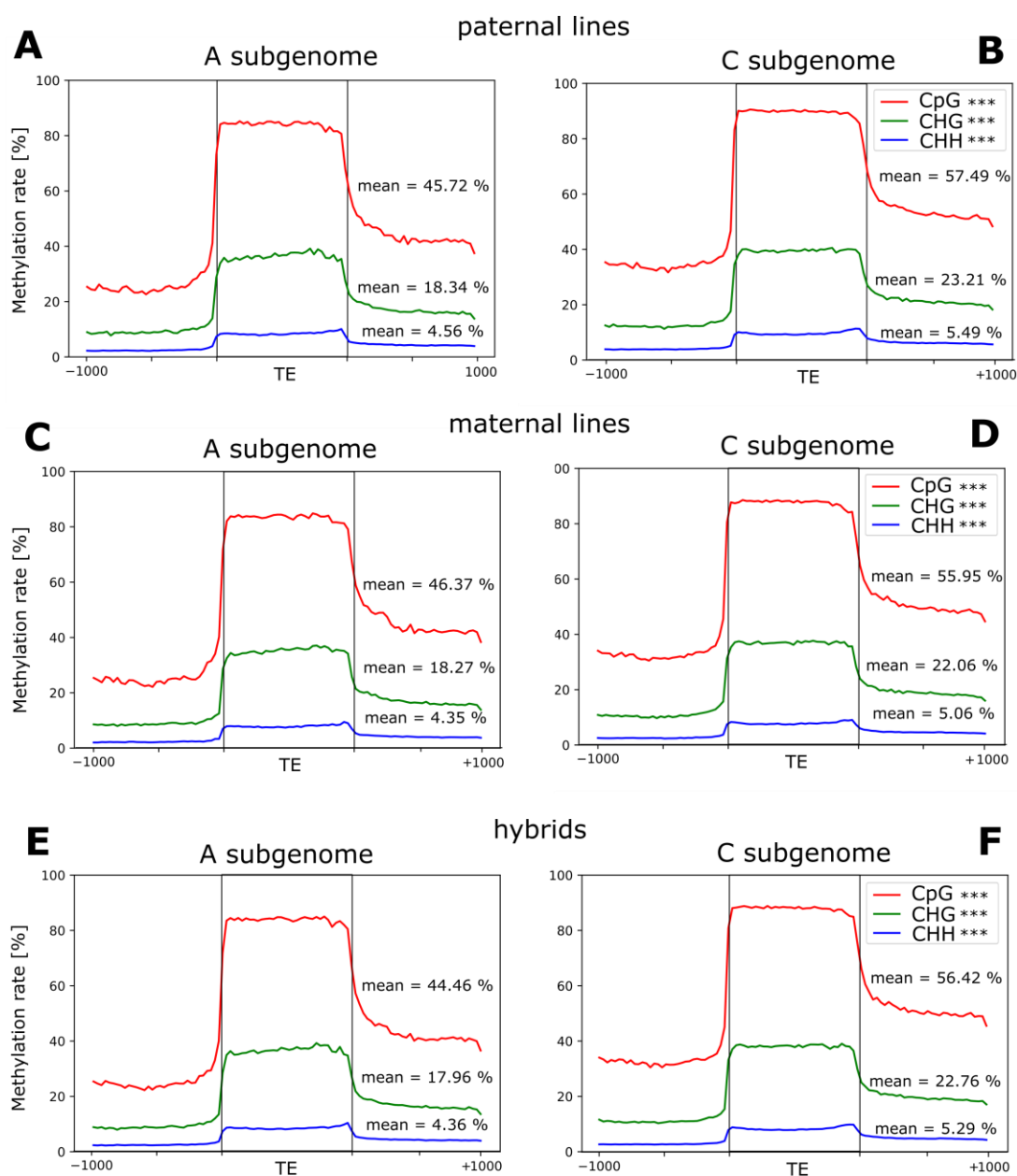


Figure 3.20: Cytosine methylation rate distribution in 100 bins over TEs including 1kb up- and downstream regions of *Brassica napus*.

Cytosine methylation rate distribution is shown for **A.** A subgenome and **B.** C subgenome of paternal lines, **C.** A subgenome and **D.** C subgenome of maternal lines and **E.** A subgenome and **F.** C subgenome of hybrid lines.

Figure 3.19 and **Figure 3.20** reveal the differences between methylation at coding sequences and transposable elements: The overall DNA methylation rate of CDS was sig-

nificantly lower than at TEs. At CG contexts, a mean methylation rate of about 14 % was found in CDS and flanking regions of the A genome, whereas at TEs, the mean was increased to ~45 %. The C genome revealed a similar pattern on a higher DNA methylation level, where the mean was increased from ~22 % to ~56 %. There were no significant differences between the paternal, maternal and hybrid group for CG methylation rates, neither in CDS nor TE, although hybrids exhibited a slight decrease in CG mean methylation rates compared to the parents at TEs. CHGs at CDS were again lower methylated than at TEs; the mean methylation rate was increasing from ~4 % to ~18 % on the A genome and from ~7 % to ~23 % on the C genome. No significant differences between the groups were visible, although again, hybrids displayed slightly lower DNA methylation rates than the parents. CHHs revealed an overall low methylation level being lowest in coding sequences with their flanking regions (~1.5 %) at the A genome and increasing at CDS at the C genome (~2 %). TEs were more methylated at CHH loci with a methylation mean at TEs located on the A genome of ~4 % and located on the C genome of ~5 %.

To analyze whether there is a direct correlation between phenotypic traits and DMRs and/or methylation mean values on features in hybrids, respectively, Pearson correlation coefficients were calculated for phenotypic and methylation traits. The heatmap **Figure 3.21** shows correlations of phenotypic traits with each other; **Figure 3.22** shows correlations first divided in a context specific manner, than in DMR number (each in total, at CDS and at TEs), relative number of TCM and TCdM events (each in total, at CDS and at TEs, respectively) the means of the DMRs and the absolute methylation of the hybrids in these regions and inheritance patterns (with additive, dominant, overdominant, maternal and paternal contribution) with seedling root/shoot traits and field traits. The seedling root and shoot traits exhibited strong positive correlations with each other ($r > 0.8$, $p < 0.01$), except the root branching angle trait and root-to-shoot ratio (**Figure 3.21**). Among the field traits, no consistent correlation was observed (**Figure 3.21**).

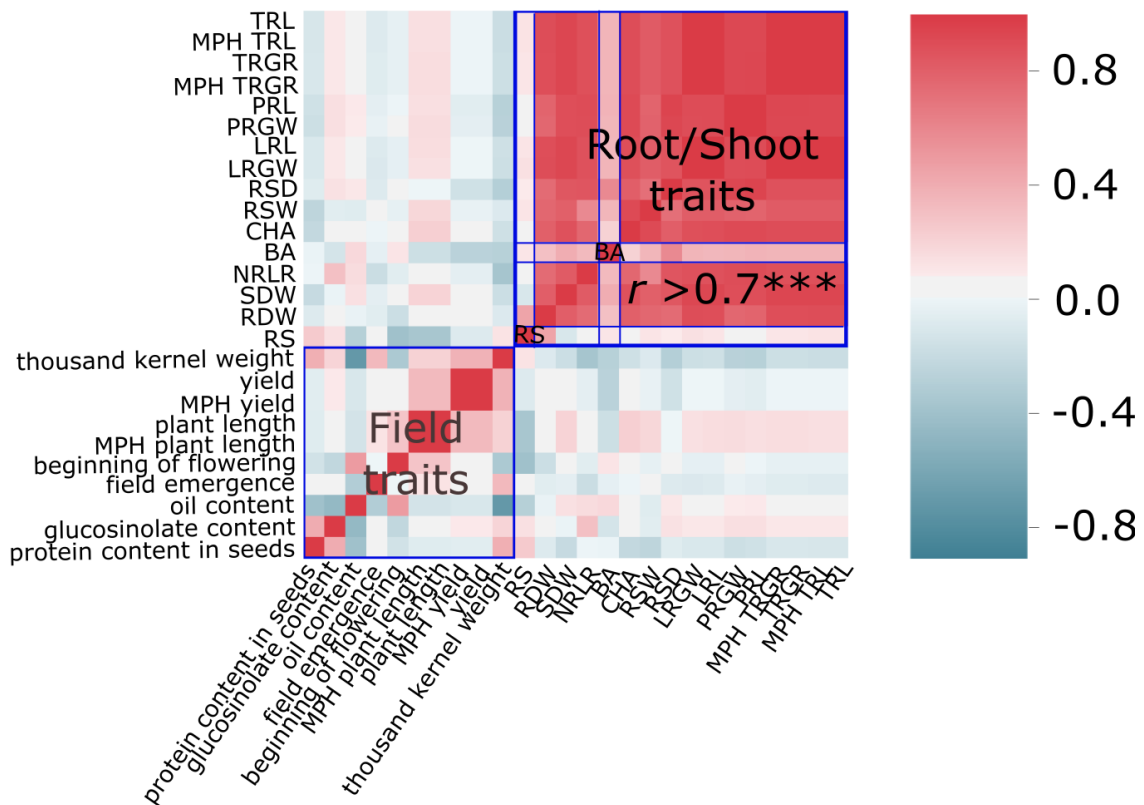


Figure 3.21: Heatmap based on Pearson correlation coefficients of field traits of mature plants and root and shoot seedling traits.

*** indicates $p < 0.01$.

Strongest correlations were revealed between DMR number (all of total, at CDS and at TEs) in any context and oil content in mature plants ($r > 0.45$ with $p < 0.05$, **Figure 3.22**). Moderate positive correlations were detected between root and shoot data of seedlings (except root branching angle) and CG tranchromosomal demethylation events on transposable elements expanding to the sum of TCM and TCdM events on TEs ($r > 0.3$). An overdominant methylation pattern in the hybrid with an apparent maternal influence in regions, which are differentially methylated between parental lines, correlated with the same seedling traits (**Figure 3.22**).

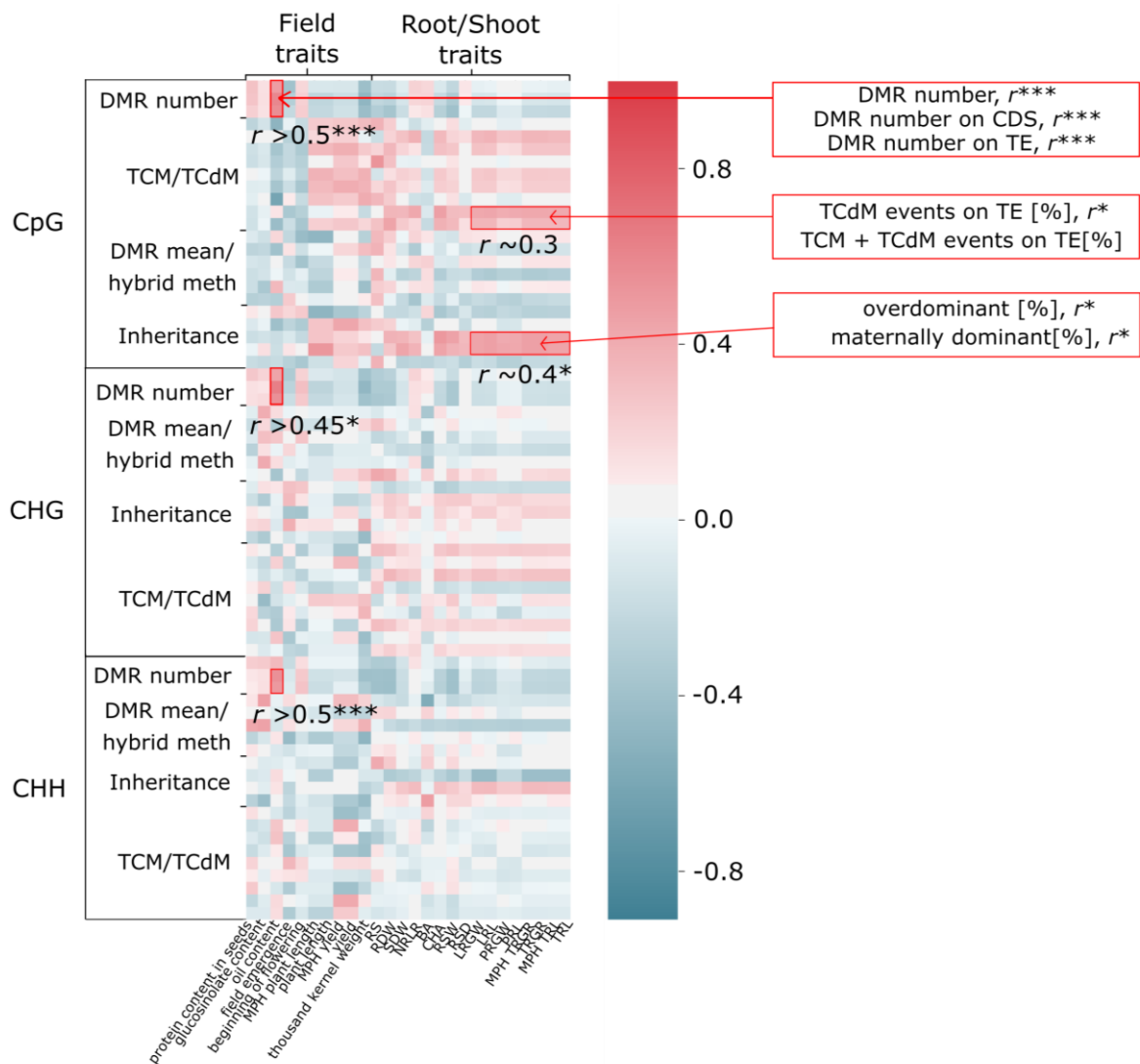


Figure 3.22: Heatmap based on Pearson coefficient shows correlations between DMR methylation means and DMR numbers (total, on CDS, on TEs) of the three contexts with field traits of mature plants and root and shoot seedling traits.

*** indicates $p < 0.01$ and * $p < 0.05$. Red rectangles indicate correlation coefficients on common level. Arrows link trait descriptions with location on the map.

3.3.5 DNA methylation and mRNA expression

DNA methylation rates of cytosine contexts from RRBS data were compared for every location covered in both, DNA methylation and mRNA expression data, to analyze the silencing effect of DNA methylation on expression of transcripts. No mRNA expression could be observed at highly methylated loci independent from the cytosine context and a high expression level was only detected, when the loci were low or nil methylated (**Figure 3.23**).

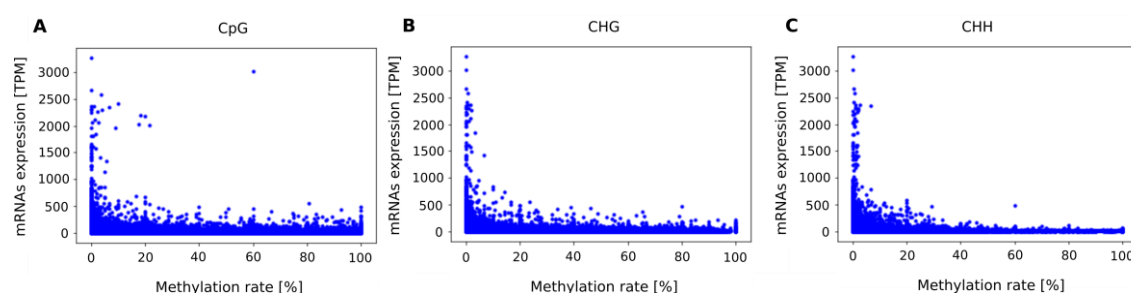


Figure 3.23: Correlation of DNA methylation rates per contexts with mRNA expression. DNA methylation rates (in %) of **A.** CG, **B.** CHG and **C.** CHH context are correlated with mRNA expression (in transcripts per million [TPM]).

Parental differentially expressed genes (DEGs) were calculated to analyze the effect on hybrid mRNA expression patterns as it was performed for DNA methylation. The fraction of additive, dominant and overdominant expression was calculated in relation to the total number of parental DEGs covered in hybrids. Hybrid values were considered as additive within a deviation of 15 % mRNA expression deviation from the parental mean. Dominant patterns were defined with 15 % mRNA expression level deviation from one of the parental values. Overdominant was defined below (for low parent) or beyond (for high parent) these values. Additive mRNA expression levels in hybrids were revealed at approximately 43 % of the loci (**Figure 3.24**). Dominant hybrid expression patterns were observed in about 33 % of the transcripts. Compared to the fraction of overdominant methylation in hybrids (**Figure 3.15**), overdominant mRNA expression patterns were more frequently occurring (>20 %).

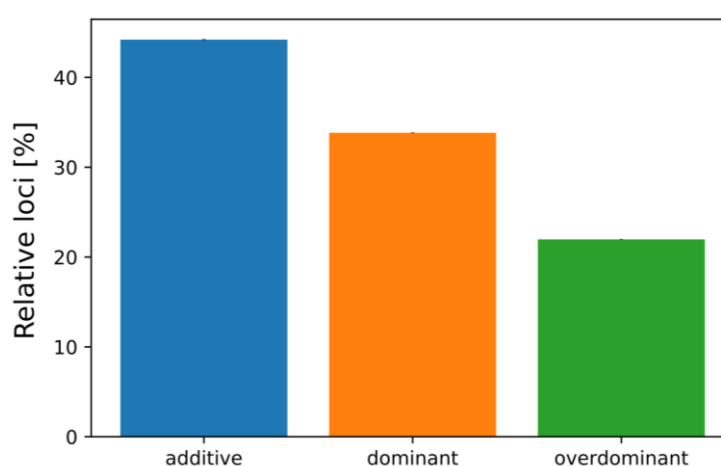


Figure 3.24: Expression patterns of transcripts of hybrids compared to parental lines relative to the total number of hybrid transcripts overlapping with parental DEGs. Black bars indicate standard deviation.

Overlapping regions of differential methylation and differential expression were identified, including 1000 bp up-and downstream, to cover also regions having an effect on expression through methylation, e.g. promoters (**Figure 3.25**). In *Brassica napus*, 747 DEGs were detected between the parents. For the CG context, from 1316 parental DMRs 11 were overlapping with the DEGs. From 1467 regions of differential CHG methylation between parents 11 showed also differential expression. The number decreased to 4 of 1026 CHH DMRs overlapping with DEGs.

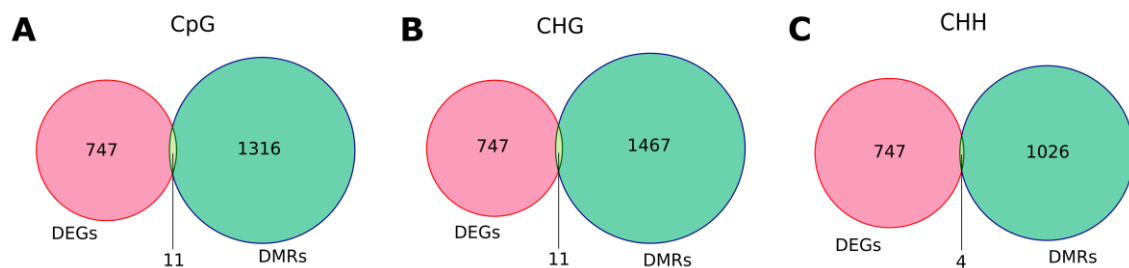


Figure 3.25: Overlap of parental DEGs with parental DMRs.

Overlap of DMRs and DEGs are shown for **A.** CpG, **B.** CHG and **C.** CHH contexts.

3.3.6 DNA methylation and sRNA expression

Small RNAs mapping to the reference genome *Brassica napus* cv. Drakkar (including ambiguously mapping ones) were characterized in distribution on annotated features (**Figure 3.26**) and size as well as expression patterns (**Figure 3.27**) to examine the connection of small RNAs and DNA methylation. To figure out, whether sRNA expression influences DNA methylation or *vice versa*, patterns of sRNA expression covered with methylation data were analyzed.

The sRNA distribution on features showed that the vast majority (>99 %) of sRNAs mapped on transposable elements (**Figure 3.26**). 20 % of the loci containing sRNA mapping sites were located in regions of parental DMRs. A small number of loci, where sRNA sequences align to, was found in non-annotated regions (~7 %) and hardly any were located on coding sequences (>1 %). sRNA mapping sites were analyzed at regions including 1000 bp up- and downstream of each feature to cover also possible promoter sequences or enhancing and inhibiting elements, respectively. This could have led to an overlap between the TEs/sRNA and the CDS/sRNA pool. Additionally, some sequences in the reference annotation files of repeats and CDS were overlapping.

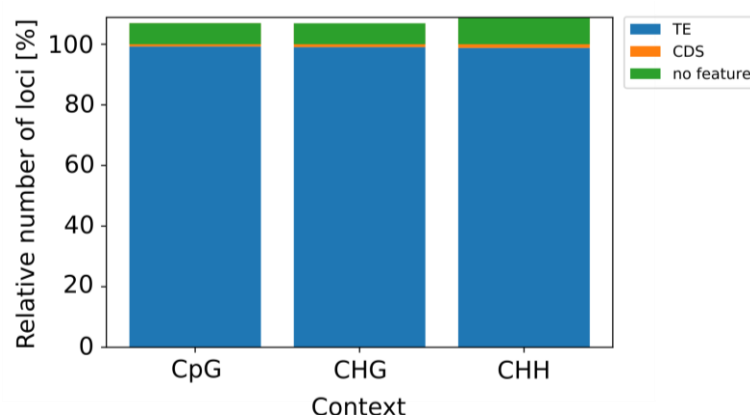


Figure 3.26: sRNA distribution on features that are exposed with cytosines covered in methylation data.

To identify the main classes of the sRNAs that map on the reference, the sequence length distribution of sRNAs was analyzed. The sequence length distribution shown in **Figure 3.27 A** revealed the range of 18 – 30 nt sRNAs in the population with a major peak at 24 nt siRNAs. The median sRNA expression showed an increase in hybrid lines compared to their parents (**Figure 3.27 B**).

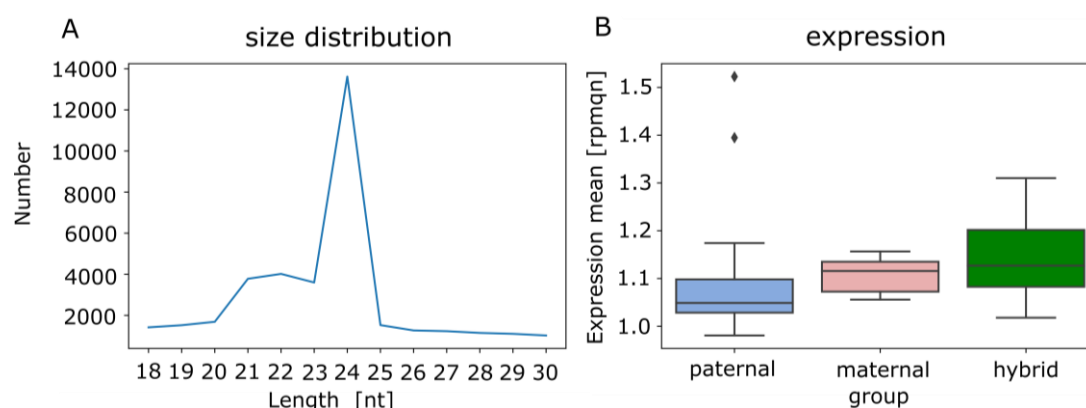


Figure 3.27: Characterization of mapped sRNAs by size and expression.

A. sRNA sequence length distribution is shown for total sRNAs. **B.** Expression patterns of total sRNAs are shown for hybrids, maternal and paternal groups. The expression is given in reads per million quantile normalized [rpmqn].

De novo methylation of all three cytosine contexts is known to be guided by 24 nt siRNAs which made up the biggest proportion of sRNAs in the samples (**Figure 3.27 A**). The effect of DNA methylation on sRNA expression and *vice versa*, respectively, was analyzed to examine how regulation is organized in *Brassica napus*. **Figure 3.28** shows the correlation of DNA methylation rate and sRNA expression at the same loci. The

pattern looked similar to the correlation of DNA methylation rates and mRNA expression (**Figure 3.23**). A decrease in sRNA expression with the increase of DNA methylation rates especially at CHH contexts in all features was observed (**Figure 3.28 A**). Less DNA methylation rates appeared together with higher sRNA expression, which was primarily present in repetitive regions (**Figure 3.28 B and C**).

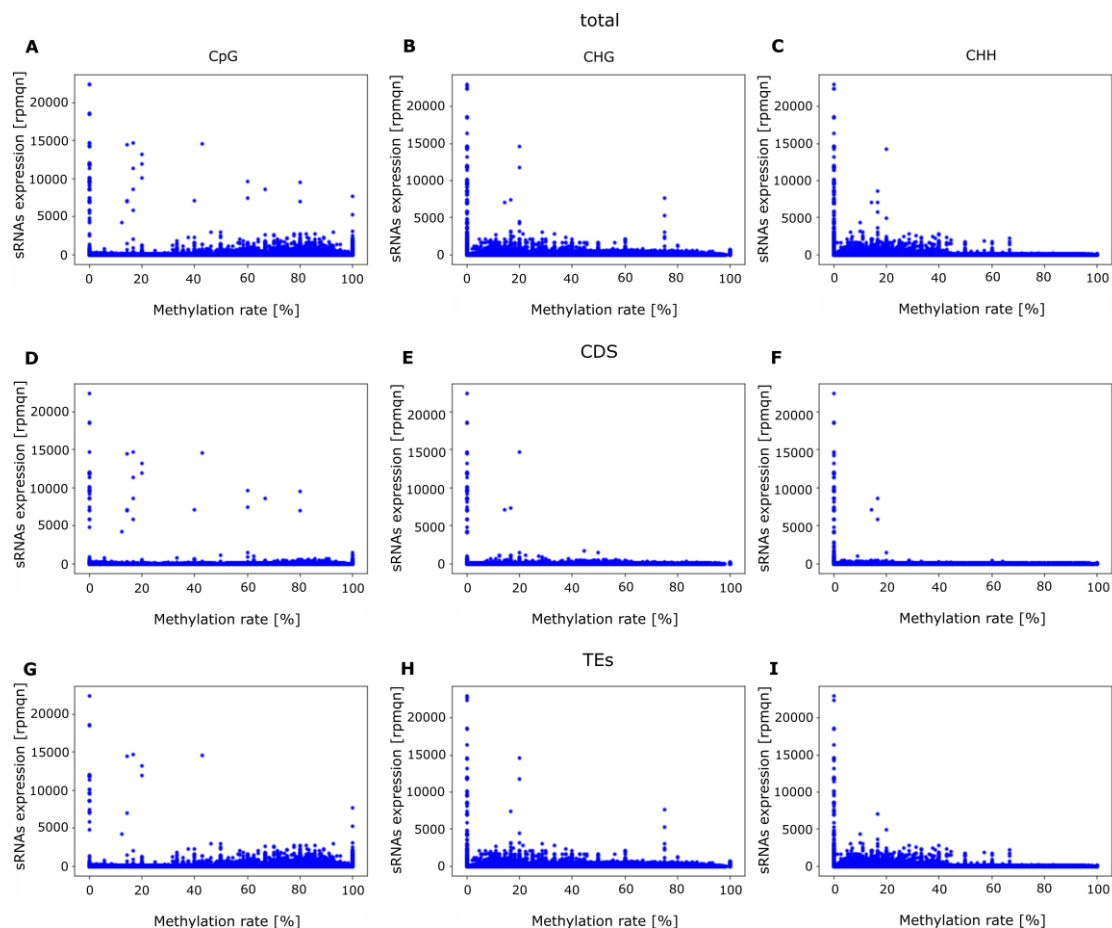


Figure 3.28: Correlation of sRNA expression with DNA methylation rates (in %) of CG, CHG and CHH contexts.

Correlations are shown for CG contexts in **A**. total, **B**. at CDS and **C**. TE loci, for CHG contexts in **D**. total, **E**. at CDS and **F**. TE loci and for CHHs context in **G**. total, **H**. at CDS and **I**. TE loci. sRNA expression is given in reads per million quantile normalized [rpmqn].

As sRNAs are gene and TE expression regulators and their expression is affected by DNA methylation and on the other hand affects DNA methylation, a direct influence on growth and growth heterosis is possible. To test this, correlation analysis of sRNA data associated with DNA methylation including DMRs and hybrid traits was performed (**Figure 3.29**) to test for a direct influence on growth and growth heterosis.

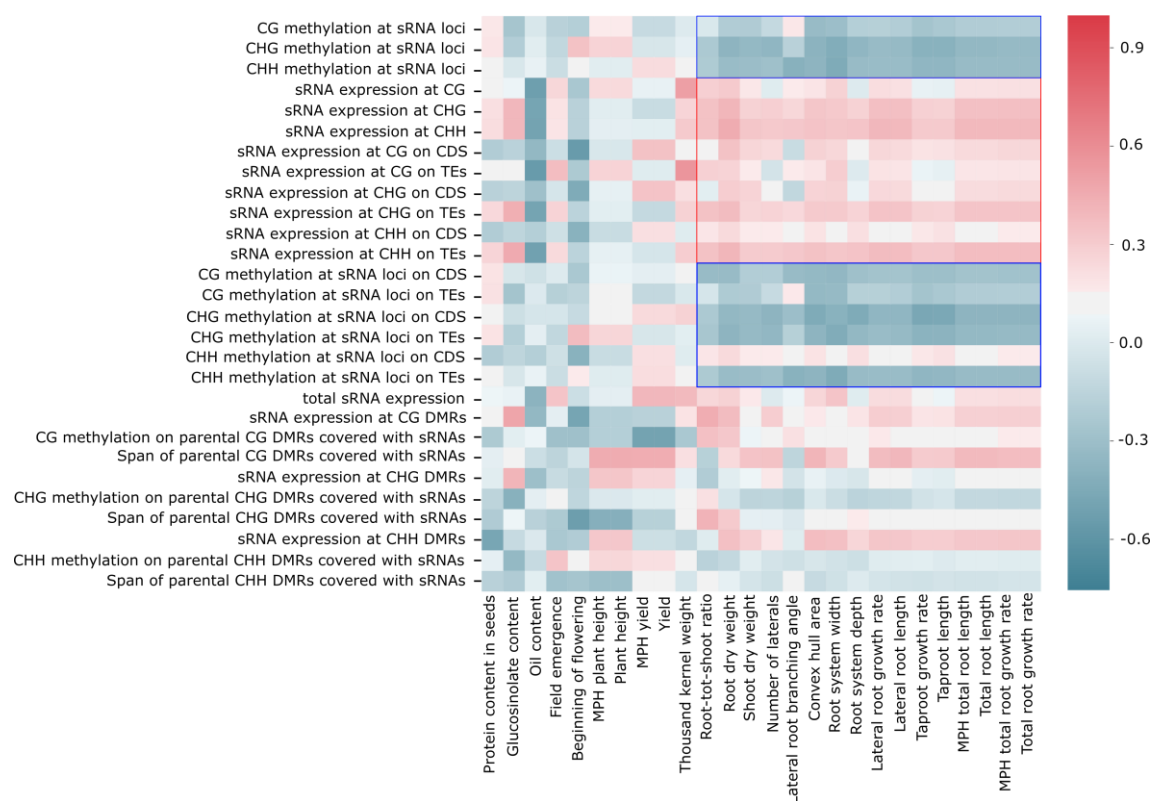


Figure 3.29: Heatmap based on Pearson correlation coefficients between sRNA expression on DNA methylation and DMR loci (total, CDS, TEs) of the three cytosine contexts with field traits of mature plants and root and shoot seedling traits of hybrids. Red rectangles indicate positive values; blue rectangles indicate negative values common for a set of phenotypic traits.

The heatmap (**Figure 3.29**) shows blocks indicating a consistent positive correlation of total sRNA expression with seedling root/shoot traits independent from cytosine context and feature (red rectangles). A negative correlation was revealed between hybrid seedling phenotypic traits and cytosine methylation rate on total sRNA loci not restricted to certain features or cytosine contexts (blue rectangle) although the correlations were weak ($-0.3 < r < 0.3$).

3.3.7 Analysis of genes involved in regulatory pathways

Since changes in DNA methylation and sRNA expression appeared between hybrids and parents, genes involved in methylation pathways and known heterotic responses, e.g. stress response via salicylic acid (SA) pathway, were tested for differences in DNA methylation and mRNA as well as sRNA expression (**Table A.3**). These included maintenance and *de novo* methyltransferases (MET1, CMT3, DRM2), genes involved in RdDM (AGO4), and the chromatin remodeler DDM1. A gene shown to differ in hybrids

compared parental inbred lines, CCA1 (*Circadian clock-associated 1*, Shen *et al.*, 2012), or genes active in SA stress response (C4H, MOS1, ICS1, NPR1, Li *et al.*, 2011; Zhang *et al.*, 2016a; Shen *et al.*, 2017) were selected to represent important pathways, which may be involved in heterosis formation. **Table 3.5** summarizes the trends of mRNA and sRNA expression as well as DNA methylation changes of these genes from parents to hybrids. DMRs and DEGs could not be detected in or close to selected gene loci. The general DNA methylation rates of the genes were mainly between 10 % and 20 % at CGs, between 1 % and 2 % at CHGs and 0.2 % and 1% at CHHs (**Table A.6**). A striking exception was MOS1, which was highly methylated in all cytosine contexts (CG: ~82 %, CHG: ~35 % and CHH: 6 %). For this gene, no mRNA expression could be detected. The mRNA expression of methyltransferases was lower than of the stress response related genes (~ 1.5-3 FPKM and 1.5-6 FPKM, respectively). In contrast to the relatively low mRNA expression levels of MET1, CMT3, DRM2 and AGO4, mRNA expression of DDM1 was about 33 FPKM, which was further increasing in hybrids to ~34 FPKM (**Table A.7**).

Table 3.5: Summary of mRNA and sRNA expression as well as DNA methylation changes of parents vs. hybrids.

Arrows indicate either up- or down regulation in mRNA/sRNA expression or DNA methylation rate of the respective genes with red color = hybrids showing significant differences compared to maternal lines, blue = significant differences to paternal lines and green = significant differences to both parents (Student's t test: $p > 0.05$). The - indicates no data detection for the gene.

		genes involved in methylation pathways						genes involved in stress response				Circadian clock
		mainte-nance		RdDM		deme-thyla-tion	C4H	MO S1	ICS 1	NP R1		
parental lines → hybrids		ME T1	CMT 3	DRM 2	AG O4						DDM 1	ROS1
Methyla-tion	CpG	↓	-	-	-	↓	↑	↑	↑	↑	↑	↑
	CHG	↓	-	-	-	≈	↑	↑	↓	↑	↓	≈
	CHH	↓	≈	-	↓	≈	↓	↓	↑	↑	↓	≈
sRNA expression		↑	↑	↓	↑	↓	-	-	↑	-	-	-
mRNA expression		≈	↓	↓	↓	↑	↑	↓	-	↓	↓	↓

A trend of down-regulation from parents to hybrids of most analyzed methyltransferases could be observed, which was significant for DRM2 (both parents to hybrids) and CMT3 (paternal lines to hybrids). AGO4, which builds a complex with 24 nt siRNAs during RdDM, exhibited a significant decrease in mRNA expression (pa-

ternal lines to hybrids). For DDM1, an up-regulation in mRNA expression was observed. The DNA glycosylase ROS1 involved in DNA demethylation and stress response revealed a higher mRNA expression in hybrids compared to their parental inbred lines, together with a higher CG and CHG, but lower CHH methylation level. Furthermore, the genes selected to represent the SA-dependent stress response pathway exhibited a general trend in down-regulation in combination with an overall slight increase in DNA methylation level. CCA1 revealed an increase in CG methylation level together with a significant decrease (paternal lines to hybrids) in mRNA expression. Pearson correlation analysis of DNA methylation rates, mRNA and sRNA expression with phenotypic traits was performed to conclude from changes in DNA methylation and RNA expression of the analyzed genes on phenotypic differences between parents and hybrids (**Figure 3.30**).

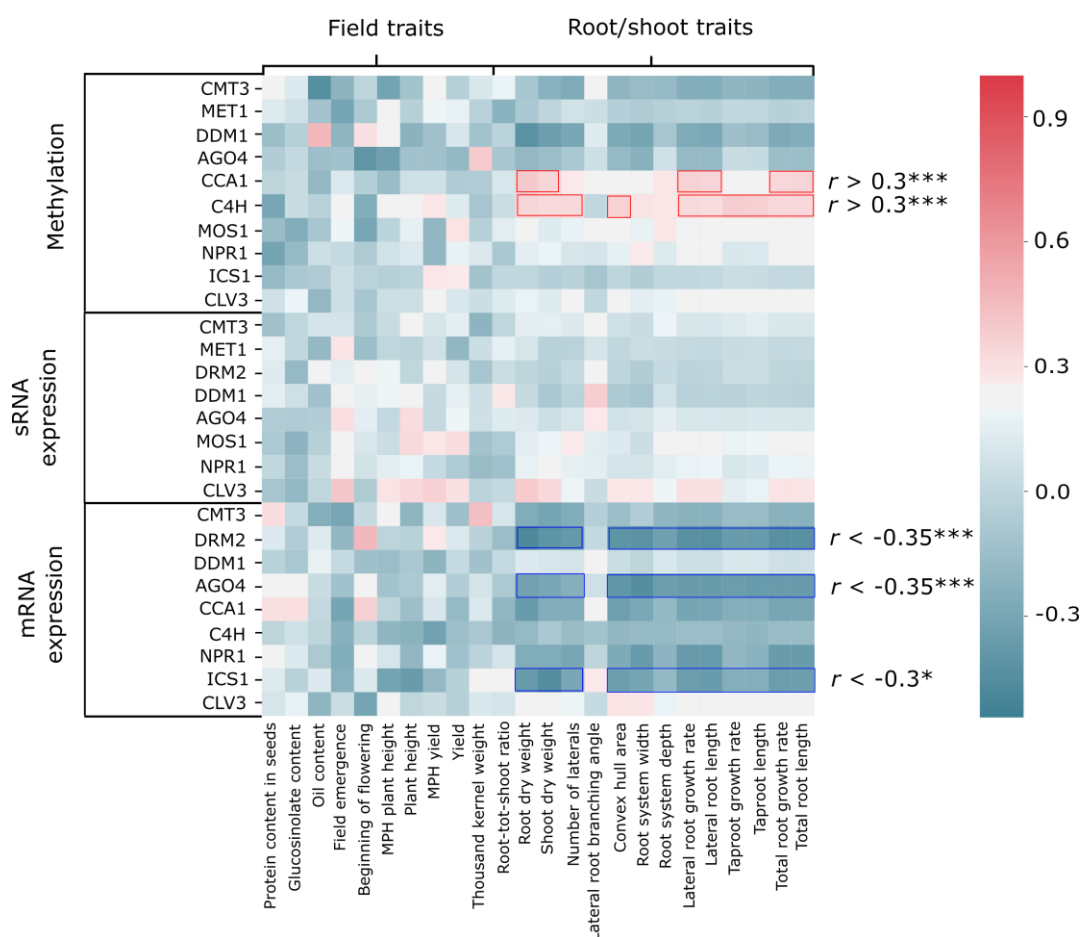


Figure 3.30: Heatmap based on Pearson correlation coefficients shown for DNA methylation rate, sRNA and mRNA expression means of selected gene loci with field traits of mature plants and root/shoot seedling traits.

Red rectangles indicate positive pearson coefficients, blue rectangles indicate negative pearson coefficient. * indicates $p < 0.05$ and *** $p < 0.01$.

The analysis revealed a positive correlation of root and shoot seedling traits with DNA methylation levels of CCA1 and C4H together with a weak negative correlation of CCA1 and C4H mRNA expression with root/shoot traits (**Figure 3.30**). Additionally, the mRNA expressions of DRM2, AGO4 and ICS1 were significantly negatively correlated with the seedling traits.

3.3.8 Summary of DNA methylation and RNA expression characteristics

General trends and differences between parents and hybrids on the DNA methylation and RNA expression level were observed in *Brassica napus*. **Figure 3.31** shows representative mRNA expression and DNA methylation traits, whose patterns were significantly changing from parents to hybrids: The expression of transcripts differing between the parents was significantly decreased in the hybrids from ~10 FPKM to ~4 FPKM. Moreover, the CHH methylation rate especially on sRNA loci found in TEs was significantly reduced in hybrids.

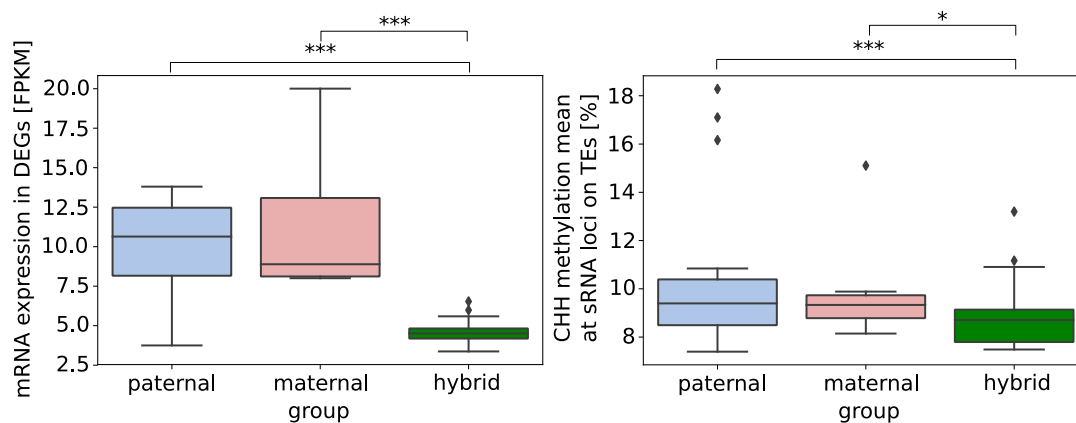


Figure 3.31: mRNA expression in DEGs and CHH methylation at sRNA locations in TEs shown as representative traits significantly differing between parents and hybrids.

*** indicates $p < 0.01$, * indicates $p < 0.05$ (Student's t test).

Although statistical significances were rare, general trends were visible (**Table 3.6**). The DNA methylation rates of every context were slightly decreasing from parents to hybrids, especially in transposable elements, where hybrids significantly differed from paternal lines in CHG and CHH contexts (**Table 3.6**). Furthermore, lower DNA methylation rates were observed at sequences, which are associated to sRNAs, together with a higher sRNA expression in hybrids. Whereas the mRNA expression in general increased in hybrids compared to parental lines, the number of transcripts, which were differen-

tially expressed between parents, was significantly reduced in hybrids (**Figure 3.31, Table 3.8**).

Table 3.6: Summary of mRNA and sRNA expression as well as DNA methylation changes of parents vs. hybrids.

Arrows indicate either up- or down regulation in sRNA/mRNA expression or DNA methylation divided by cytosine context with red color = hybrids showing significant differences compared to maternal lines, blue = significant differences to paternal lines and green = significant differences to both parents (Student's t test: $p > 0.05$).

parental lines → hybrids		total		CDS		TE	
Methylation	CpG	↓		↓		↓	
	CHG	↓		↓		↓	
	CHH	↓		≈		↓	
		methylation	expression	methylation	expression	methylation	expression
sRNA ↑	CpG	≈	↑	↓	↑	≈	↑
	CHG	↓	↑	↓	↑	↓	↑
	CHH	↓	↑	↑	↑	≈	↑
mRNA expression in parental DEGs		↓***					
mRNA expression total		↑					

To examine the effect of DNA methylation as well as sRNA and mRNA expression patterns on the level of heterosis, the hybrids were grouped in low and high performing hybrids (each 50 %) by their MPH value of total root growth rate in seedlings, which was chosen as a highly heterotic trait. As from parents to hybrids, total sRNA expression was significantly increasing comparing low with high heterotic hybrids (**Figure 3.32**).

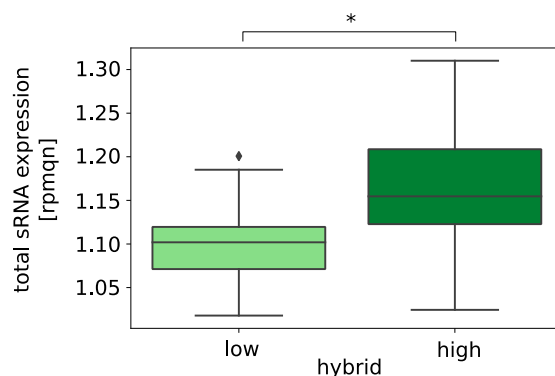


Figure 3.32: sRNA expression levels shown as representative traits significantly differ between low performing and high performing hybrids.

* indicates $p < 0.05$ (Student's t test).

Trends of changes in DNA methylation and RNA expression patterns between low and high heterotic hybrids were less pronounced as trends observed from parents to hybrids (**Table 3.7**). Both, the number of parental DMRs of all cytosine contexts for each hybrid and CG methylation rates in these regions decreased with higher MPH independent of the annotation feature CDS or TE. On the other hand, the CHG and CHH methylation rates of TEs in parental DMRs in hybrids increased with the MPH value. The DNA methylation rates in general were not showing a constant overall pattern (**Table 3.7**). Expression differences of sRNAs were heterogeneously distributed and seemed to depend on cytosine context and/or annotation feature. In parental DEGs, DNA methylation rates and mRNA expression exhibited variable patterns over the different cytosine contexts. In contrast, the DNA methylation rates at mRNA transcribing regions in total showed a trend of increasing from low to high heterotic hybrids leading to the observed decrease of transcripts (**Table 3.7**).

Table 3.7: Summary of mRNA and sRNA expression as well as DNA methylation changes divided by cytosine context of low performing hybrids vs. high performing hybrids in the phenotypic trait MPH of seedling total root growth rate.

Arrows indicate either up- or down regulation in mRNA or sRNA expression or DNA methylation at the respective context with green = significant differences (Student's t test: $p > 0.05$).

MPH total root growth rate low hybrids → high hybrids		total				CDS				TE			
		DMRs		total	DMRs		total	DMRs		total			
		nr	meth		nr	meth		nr	meth				
Methylation	CpG	↓	↓	↑		↓	↓	↓		↓	↓	↑	
	CHG	↓	↑	↑		↓	≈	↓		↓	↑	↑	
	CHH	↓	↑	↑		↓	↑	↓		↓	↑	↓	
		expr	meth	expr	meth	expr	meth	expr	meth	expr	meth	expr	meth
sRNA ↑	CpG	↑	≈	↓	↓	↓	↓	↑	↓	↓	↓	↑	↓
	CHG	↓	↑	↑	↓	↑	↓	↑	↓	↓	≈	↑	↓
	CHH	↑	↑	↓	↓	↓	↑	↑	↑	↓	↑	↑	↓
		DEG		total									
		expr	meth	expr	meth								
mRNA	CpG	↓	↓	↓	↑								
	CHG	↑	↑	↓	↑								
	CHH	↑	↓	↓	≈								
	total	↑		↓									

Hybrid DNA methylation patterns were analyzed in regions, where parental differential methylation was observed, to investigate DNA methylation variation emerging with increasing MPH of total root growth rate. **Table 3.8** summarizes alterations between

low and high heterotic hybrids: Additive DNA methylation patterns were reduced in every cytosine context from low to high heterotic hybrids. In contrast, dominant and overdominant methylated loci increased from low to high heterotic hybrids. In high heterotic hybrids, the maternal dominance in CG methylation patterning in hybrids significantly increased together with a decreasing paternal dominance compared to low heterotic hybrids. The opposite was observed for the CHH context. TCM and TCdM events on a genome-wide manner both increased at the same time from low to high heterotic hybrids, which was true for all cytosine contexts and due to the observed increase of both events in general. This in turn was observed together with a declining number of additive DNA methylation patterns from low to high heterotic hybrids, which indicated a trend of increasing dominant DNA methylation patterns combined with decreasing additive DNA methylation patterns from low to high heterotic hybrids. Regions located in or close to CDS however, revealed less TCM events from low to high heterotic hybrids, although here, too, the relative number of both, TCM and TCdM events, increased.

Table 3.8: Summary of hybrid DNA methylation changes in regions of parental DMRs of low performing hybrids vs. high performing hybrids in the phenotypic trait MPH of seedling total root growth rate.

Arrows indicate either decrease or increase of DNA methylation rate at the respective context with green = significant differences (Student's t test: $p > 0.05$).

MPH total root growth rate low hybrids → high hybrids		Methylated context		
		CpG	CHG	CHH
Additive loci		↓	↓	↓
dominant loci		↑	↑	↓
overdominant loci		↑	↑	↑
maternally dominant loci		↑	≈	↓
paternally dominant loci		↓	↑	↑
total	TCM events	↑	↑	↑
	TCdM events	↑	↑	↑
	TCM + TCdM events	↑	↑	↑
CDS	TCM events	≈	↓	↓
	TCdM events	↑	↑	↑
	TCM + TCdM events	↑	↑	↑
TE	TCM events	↑	≈	↑
	TCdM events	↑	≈	↓
	TCM + TCdM events	↑	↑	↑

3.3.9 Heterosis-associated DMRs

Parental DMRs were associated to MPH of total root growth rate using an approach based on binomial distribution (Seifert *et al.* 2018, 2.2.3.3) to analyze, whether specific parental differentially methylated regions may be involved in the increase of heterosis in hybrids. Parental DMRs were discovered, which were mainly negatively associated to heterosis (**Table A.8**). The CG context revealed 15 DMRs to be associated with MPH of total root growth rate, from which two were more often found in high heterotic

hybrids (= positive ha-DMR). Seven parental CHG DMRs and one CHH DMR were exclusively negatively associated to MPH of total root growth rate.

The correlation of the ha-DMRs with DNA methylation rates of these regions in hybrids was analyzed. The difference within the ha-DMRs (ΔmC) as well as their absolute number was correlated to total root growth rate MPH of seedlings. Expressions of sRNAs collocated with the DMR were determined to investigate possible effects of RdDM. Neither CG methylation rates in hybrids (**Figure 3.33 A**) nor ΔmCG of positive or negative ha-CG DMRs (**Figure 3.33 B**) showed a correlation with MPH of total root growth rate. The number of negatively ha-DMRs was significantly decreased with rising MPH (**Figure 3.33 C**); less ha-DMRs were present in the hybrids with higher MPH. A slight increase in number of positively ha-DMRs was observed with increasing MPH (**Figure 3.33 C**). sRNA expression in hybrids at these loci did not reveal a specific pattern (**Figure 3.33 D**). Furthermore, for ha-CHG DMRs none of the features mentioned before (**Figure 3.34 A-D**) significantly correlated with MPH.

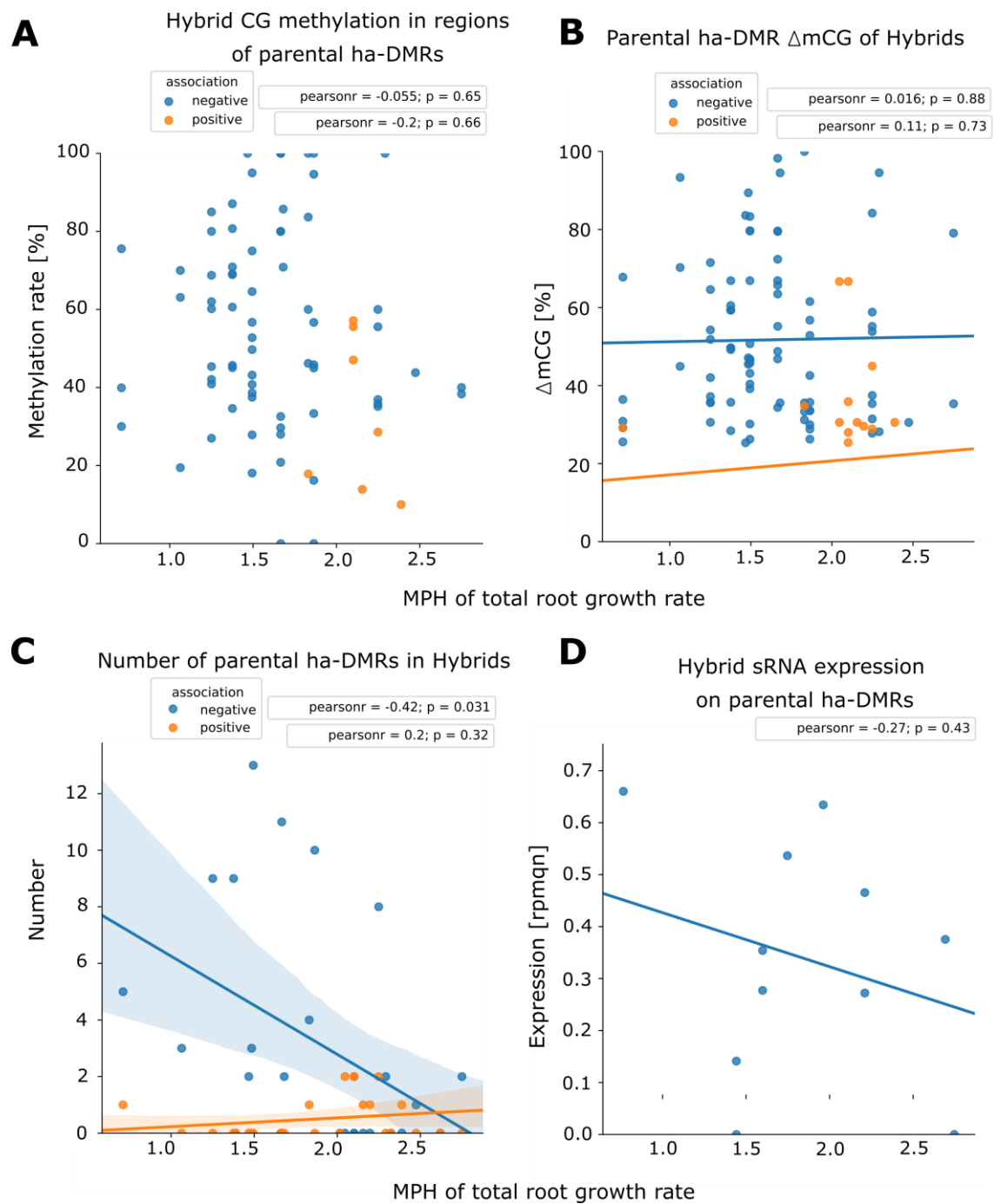


Figure 3.33: Characterization of heterosis-associated CG DMRs and correlated features.

Ha-CG DMRs correlated with the **A**. CG methylation rate of hybrids in the corresponding region, **B**. the difference within the DMR (Δ mCG, divided into positively and negatively associated DMRs), **C**. the absolute number of DMRs for the hybrids and **D**. the expression of sRNAs located in the corresponding regions in the hybrid. Pearson correlation coefficient and corresponding probability are given in grey rectangles.

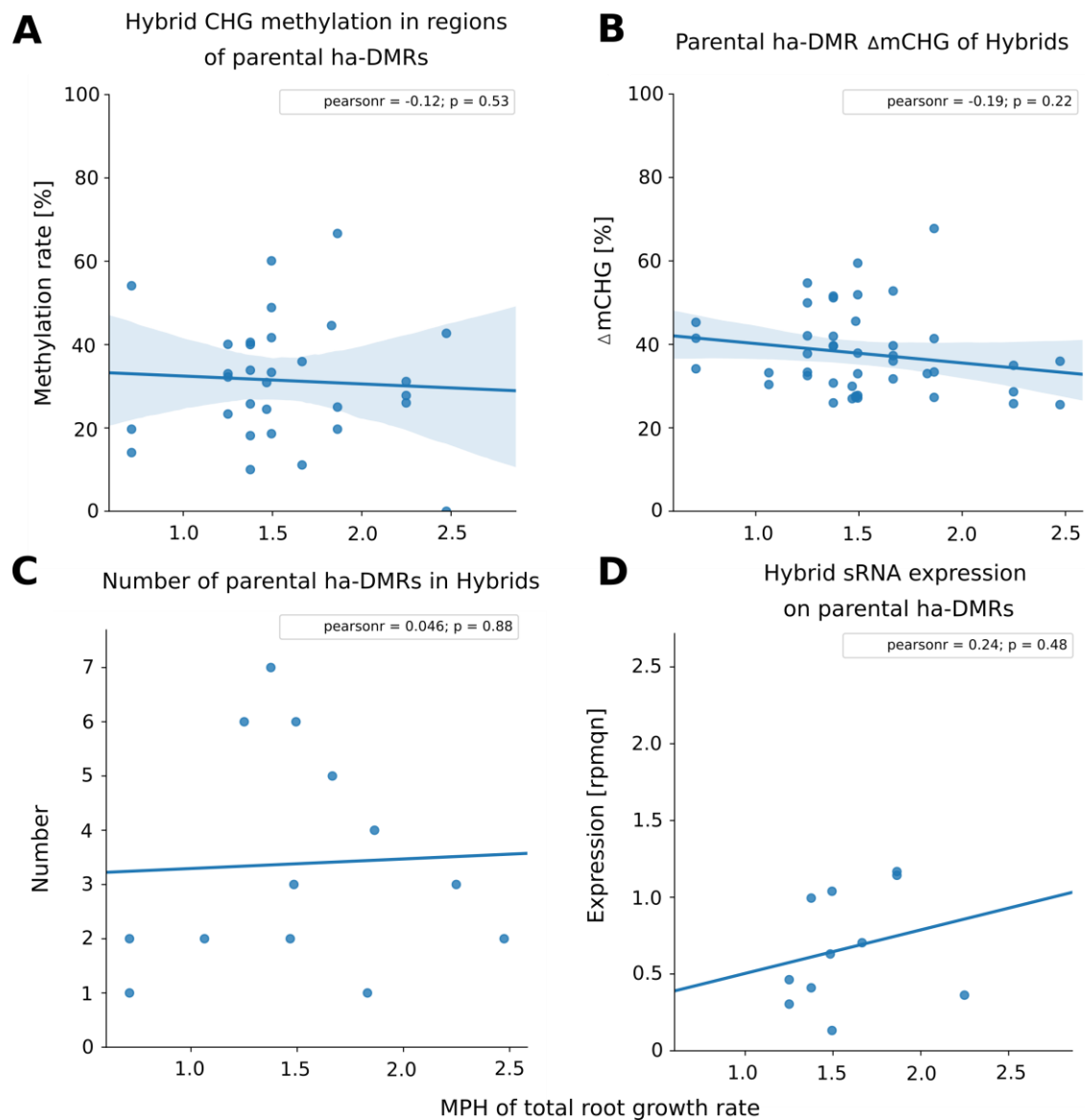


Figure 3.34: Characterization of heterosis-associated CHG DMRs and correlated features.

Ha-CHG DMRs correlated with the **A.** CHG methylation rate of hybrids in the corresponding region, **B.** the difference within the DMR (Δ mCHG), **C.** the absolute number of DMRs for the hybrids and **D.** the expression of sRNAs located in the corresponding regions in the hybrid. Pearson correlation coefficient and corresponding probability are given in grey rectangles.

4 Discussion

4.1 Reduced representation bisulfite sequencing method

Genome-wide profiling of plant methylomes has been useful to study phenotypic variations (Kawakatsu *et al.*, 2016), adaptation to environmental stresses (reviewed in Elhamamsy, 2016) or during development (Zhang *et al.*, 2010). Monitoring methylation in multiple samples or of large genomes on single base resolution is still a considerable cost factor due to library preparation and sequencing. Reduced representation bisulfite sequencing (RRBS) provides a cost-effective alternative compared to whole genome bisulfite sequencing (WGBS), as it allows comparison of the randomly selected regions over a range of samples with less need of sequencing depth (Bock *et al.*, 2010). To find the most suitable approach for the used genomes and sample sizes, RRBS protocols were varied.

For genomic DNA (gDNA) extraction, the phenol/chloroform protocol modified after the Waite lab (Pallotta *et al.*, 2000) was used, since it was proven in the group to be a high yielding gDNA extraction method for various maize tissues. The next step that was examined for potential improvement was the restriction with endonucleases. Several restriction enzymes and combinations were tested for fragment size and distribution. The gDNA digestion with MspI and ApeKI revealed less fragment sizes in a range of 300 to 700 bp, whereas Taq^αI gDNA restriction enriched fragments evenly with both fragment size and concentration increasing. This pattern emerged to be suitable with regard to later bisulfite conversion and sequencing in terms of additional fragmentation and sequence yield, therefore it was used for all following RRBS library preparations. Common RRBS protocols using MspI rather enrich CG-rich sequences in the genome, which is suitable for mammalian genomes (Andrews, 2013). It enables the majority of CG islands, promoters or other relevant genomic regions to be captured with limited sequence data (Gu *et al.*, 2011). However, plants have a different pattern of CG distribution and methylation across the genome and lack of highly methylated CG islands (Feng *et al.*, 2010). As the restriction site of Taq^αI does not cut a cytosine context (like MspI), no methylation information will be lost. On the other hand, since CG is con-

tained in the recognized restriction site, this might lead to favoring sequences containing CGs, but not in an excessive manner compared to MspI.

A second step to select regions over the genome and increase comparability over multiple samples is the subsequent size selection. A cost-efficient custom column (2.2.2.1 a) was used without a great loss of DNA. As additional fragmentation takes place during bisulfite conversion (Kint *et al.*, 2018), larger fragments (1000-1200 bp) were chosen compared to common RRBS protocols (40-220 bp, Andrews, 2013). Taken together restriction and size selection, estimation of potential genome and cytosine coverage via *in silico* restriction showed a suitable pattern of a combination of Taq^αI with a fragment size of 1000-1200 bp, which was about 9 % for *Brassica napus* and 13 % for *Zea mays*, respectively (compared to 14 % in Mager, 2018). However, the conversion fragmentation effect still represents an unknown variable, although the values base on the assumption that additional fragmentation to 40-220 bp occurs. Taq^αI *in silico* digestion produced sufficient potential genome and cytosine coverage for this fragment size, too, whereas MspI and ApeKI restrictions revealed very low potential coverages for *Brassica napus* reference genome restriction, which could be insufficient for sequencing libraries. Another problem with accuracy of the *in silico* restriction might be SNP-based variation of used genotypes deviating from the reference genome, which can likely cause significant deviations from predicted results, as genetic diversity between genotypes of different cultivars and lines was observed for *Brassica napus* by several studies (Halldén *et al.*, 1994; Becker H *et al.*, 1995; Shengwu *et al.*, 2003).

For the actual bisulfite sequencing library preparation, two methods were tested: The first protocol was based on the NextFlex Kit by Bioo Scientific. The second method was adapted from a protocol published by Smallwood *et al.* (2014). Despite the advantages of the method developed by Smallwood *et al.* (2014) itself, which include less DNA input amount, post-bisulfite adapter tagging (PBAT) resulting in a greater amount of tagged fragments and lower cost per sample (~5 EUR), the sequencing results uncovered more advantages of this method: The libraries prepared with the NextFlex protocol showed an abnormal high overall methylation rate of all contexts compared with the literature of *Brassica rapa* (Chen *et al.*, 2015; Niederhuth *et al.*, 2016), *Brassica oleracea* (Niederhuth *et al.*, 2016) and *Brassica napus* (Shen *et al.*, 2017) and with the Smallwood libraries (**Table A.9**). These results preferred usage of the protocol adapted from Smallwood *et al.* (2014).

The obtained sequencing reads were processed via trimming, mapping and calculation of bisulfite conversion rate. The reads were mapped to the respective reference genome (*Zea mays* or *Brassica napus*, 2.2.3). The mapping of maize sequencing reads resulted in a relatively low mapping efficiency. This could be due to high amount of repetitive regions in the maize genome (>80 %, Schnable *et al.*, 2012). As only uniquely mapping reads were included in the analysis, some of the repetitive regions covered in the data could be excluded due to ambiguous mapping. This is supported by the relatively higher amount of ambiguously mapping reads of the maize data compared to the rapeseed data probably resulting from the smaller *Brassica napus* genome containing less transposable elements (35 %, Chalhoub *et al.*, 2014). Moreover, all the libraries sequenced could also exhibit low mapping efficiency, because the used genotypes may differ from the reference (Halldén *et al.*, 1994; Becker H *et al.*, 1995; Shengwu *et al.*, 2003). This further explains the differences of mapping efficiencies between the genotypes of the *Brassica napus* population and was the reason to lower the minimum score threshold for valid alignments in Bismark.

In conclusion, the reduction of represented genomic regions lowered the sequencing cost considerably while providing assessment of total genome-wide DNA methylation levels and assessment of DNA methylation in categorical genomic regions. However, overlapping cytosines between all samples were rare. All in all, a RRBS protocol has been developed allowing comparative analysis of multiple or large genome plant samples to detect methylation levels on single-nucleotide resolution and providing quantitative DNA methylation measurements at considerably lower cost compared to WGBS.

4.2 Impact of DNA methylation on heterosis in maize and oilseed rape

4.2.1 Early embryo maize tissue culture

Angiosperm reproductive development is highly dynamic with respect to epigenomic information (Calarco *et al.*, 2012; Yang *et al.*, 2015). During the establishment of the male germline and in gametogenesis, dramatic expression changes have been identified in many genes involved in epigenetic regulation (Borges *et al.*, 2008; Grant-Downton *et al.*, 2009). These result in unprecedented changes in the silencing and expression of repetitive sequences such as transposons as well as coding genes (Engel *et al.*, 2003; Slotkin *et al.*, 2009; Abdelsamad & Pecinka, 2014). Similarly, dynamic changes have been identified in female gametogenesis (Chettoor *et al.*, 2014). Fertilization

events therefore combine gametes, which have acquired different epigenomic states (Park *et al.*, 2016). From examples of imprinting in embryonic and endosperm tissue, it is clear that differential epigenetic states have profound post-fertilization impacts on development (Dickinson & Scholten, 2013; Gehring, 2013). The impact of combining gametes from genetically different parents, such as diverged inbred parental lines, is striking as it may result in heterosis. DNA methylation in this immediate post-fertilization developmental context could therefore contribute to establishing heterotic phenotypes. An *in vitro* system using maize lines to permit investigation of cytosine methylation in early embryogenesis was developed. This involved the isolation and culture of one-day-old embryos, which enabled study of subsequent growth traits in highly controlled environments. By applying the pharmacological agent aza that efficiently blocks DNA methylation (Griffin *et al.*, 2016), artificial demethylation was induced during early embryogenic development *in vitro*.

The *in vitro* system proved to be an effective way to deliver aza as was shown by RRBS analysis: Hypomethylation is confirmed genome wide and in every cytosine context by the results of RRBS of 7-days-old embryos and seedlings. Methylation rate means in control samples approximate literature values with CpG (65%), CHG (50%) and CHH (5%) methylation (Regulski *et al.*, 2013). The decrease in DNA methylation artificially established early in embryogenesis is maintained during development, which indicates the strong influence of early changes on later developmental stages.

Growth rate was measured as heterotic trait to compare heterosis of non-treated and hypomethylated seedlings. In comparison to the aza-treated hybrids, the level of heterosis in the non-treated hybrids was considerably reduced the nucellus tissue culture *in vitro* with a mid-parental heterosis value of about 9 % and did not exceed the best parent value. For A188, several quantitative trait loci (QTLs) for good regeneration ability in tissue culture are known (Salvo *et al.*, 2018). As the hybrids were generated by crossing A188 with H99, just one allele with these QTLs is left which could lead to a lower regeneration rate in tissue culture. Moreover maternal nutrient supply may play a role, as it has been shown that the transcriptome profile of F1 hybrids is similar to the maternal line in the first days of development (Alonso-Peral *et al.*, 2017). At 0 days after sowing (DAS), a small number of non-additively expressed genes overlapped between reciprocal hybrids, while more non-additively expressed genes overlapped between reciprocal hybrids at 3 DAS, suggesting a weakening of the maternal effect, which is established in developing embryos from 0 to 3 DAS (Zhu *et al.*, 2016; Alonso-

Peral *et al.*, 2017). Additionally, endosperm-specific maternally expressed genes (MEGs) are involved in nutrient uptake and supply as well as in the auxin signaling pathway, concordant with the onset of starch and storage protein accumulation (Xin *et al.*, 2013). In the concerned crosses, A188 served as maternal line. Removing the embryo in the zygote stage and thereby cancel the maternal supply might lead to unexpected effects especially in the hybrids. This possible maternal effect might also cause the decreased growth rate in the non-treated reciprocal hybrid H99 x A188.

When the growth dynamics of untreated control seedlings were compared with seedlings of the same genotypes derived from 50 μM aza-treatment, striking changes were apparent. Whilst there was no significant difference between the hybrids from aza-treated and control groups, the inbred lines showed reduced growth upon aza-treatment, most probably caused by toxicity of aza (Palii *et al.*, 2008), resulting in increased heterosis of aza-treated hybrids. As a greater resistance to demethylation in the hybrids is not consistent with aza's mode of action, nor it is consistent with the present bisulfite sequencing data, a possible explanation could be that inducing demethylation during early embryogenesis promotes heterosis. This in turn compensates for any negative impacts on growth from aza-treatment and toxicity. In contrast, in *Arabidopsis* hybrids and parental lines growth reduction in all genotypes following aza-treatment was shown (Shen *et al.*, 2012). However, aza-treatment to promote demethylation in inbred *Scabiosa columbaria* lines reduced the negative effects of inbreeding depression (Vergeer *et al.*, 2012), which is in line with the results of this work. Nevertheless, both studies used aza-treatment of plant material after embryogenesis was completed. However, studies focusing on the establishment of epigenetic variation in embryogenesis furthermore partially reported an increased hybrid vigor: Lauss *et al.* generated *Arabidopsis* epiHybrids by crossing Col wildtype maternal plants to *ddm-1* derived epiRILs. They found high-parent heterosis for different quantitative traits (e.g. leaf area, main stem branching and plant height) induced by the decreased methylation rates in all three cytosine contexts which reduction of DDM1 is mediating (Lauss *et al.*, 2018). Dapp *et al.*, who used a similar approach by crossing a *met1*-derived epiRIL with Col wildtype, and observed heterosis for biomass in one of the epiHybrids (Dapp *et al.*, 2015). Importantly, both *epiRILs* and the aza-treatment act on DNA methylation pattern variation in embryogenesis.

Methylation sensitive amplification polymorphism (MSAP) sequencing results of Zhao *et al.* (2007), who found aberrant methylation patterns in seedling leaves of maize hy-

brids compared to their parental lines, and Liu *et al.* (2014), who studied seedlings, embryos and endosperm of maize hybrids and their parental lines, showed lower relative total methylation levels in the hybrids than in their corresponding mid-parent values, thus exhibited the same tendencies as shown in this work.

The study by Seifert *et al.* (2018) revealed a negative correlation between parental sRNA expression variation with grain yield heterosis in maize. This means that a majority of sRNAs associated with heterosis (ha-sRNAs) and differentially expressed in the parents of two different heterotic pools show negative correlation with grain yield heterosis (Seifert *et al.*, 2018). This is in line with an increased heterosis, which might be caused by DNA demethylation in the early embryo maize tissue culture. Methylation which might be mediated by the negative ha-sRNAs could be abolished through aza and therefore enhance heterosis.

In contrast, *Arabidopsis* hybrids displaying heterosis showed altered 24 nt siRNA levels and increased DNA methylation in post-germination plants (Greaves *et al.*, 2012; Shen *et al.*, 2012). The greater methylation observed in *Arabidopsis* and rice hybrids (He *et al.*, 2010) as well as recombinant inbred lines of maize (Regulski *et al.*, 2013) and that no effect on hybrid vigour could be detected in mutants with reduced methylation levels (Kawanabe *et al.*, 2016, Zhang *et al.*, 2016) may be due to differential inputs of RdDM-triggering sRNAs from parental genomes leading to trans-chromosomal methylation (Greaves *et al.*, 2014). Collectively, these data are consistent with hybrid formation triggering novel patterns of DNA methylation that are linked to the repression of gene expression activity at a fraction of loci.

In maize, it has been shown that in early embryogenesis CHG methylation is generally decreased which leads to transcriptional activity whereas CG methylation present in gene bodies also results in activation of expression (Lu *et al.*, 2015). Linked to hybrid performance, an enhanced gene expression activity was found in maternal alleles of *Arabidopsis* hybrids in the early embryogenesis despite the 95 % of gene expression being at the mid-parent value (Alonso-Peral *et al.*, 2017). As this gene expression decreases in the further development (Zhu *et al.*, 2016; Alonso-Peral *et al.*, 2017), this might be a hint, that an decreased DNA methylation (especially in the CHG context) leading to an increased transcription during early embryogenesis in hybrids plays a role in heterosis formation.

In *Arabidopsis thaliana*, studies showed that methylation present on TEs suppressed the expression of neighboring genes (Hollister & Gaut, 2009). It was suggested that

plants keep a dynamic balance (a trade-off effect) between gene expression and the activity of adjacent TEs (Hollister & Gaut, 2009). This balance may be mediated by 24 nt small RNAs guiding TE methylation through RdDM (Matzke & Moshier, 2014), which was found to be negatively correlated with gene expression (Cheng *et al.*, 2016). Aza likely also abolishes methylation of cytosine contexts on TEs as demethylation is globally induced by depletion of soluble methyltransferase levels (Palii *et al.*, 2008). This could lead to an activation of nearby genes possibly resulting in enhanced vigor. Since such a reduction in TE regulation by demethylation can cause genome instability (Maxwell *et al.*, 2011), this might not be a favorable effect for the plant and therefore argues against the theory that heterosis is based on the enhancement of beneficial factors. But Bar-Zvi *et al.* (2017) proposed a model that centers the impairment of growth limiting factors. In support of this, a study in budding yeast hybrids showed multiple signs of impaired regulation of growth-related pathways. This included loss of programmed cell-cycle delays, weakened repression of respiratory metabolism and altered slowdown of growth during stress (Herbst *et al.*, 2017). As aza affects also cell cycle regulation (Palii *et al.*, 2008), it might be involved in impairment of growth limiting factors in hybrids resulting in enhanced heterosis.

The increase of growth heterosis by inducing demethylation suggests that an early establishment of DNA methylation may contribute more to restraining heterosis than promoting it, primarily by repression of transcriptional activity, and concomitant reduction in protein products. The artificial DNA demethylation in early embryos may enhance natural hybrid demethylation mechanisms to increase gene expression that occur during embryo development and thus contribute to the generation of heterosis.

4.2.2 Population DNA methylation dynamics of *Brassica napus*

Epigenetic factors, such as DNA methylation or chromatin modifications, can cause phenotypic variation and contribute to adaptation in plants (Kawakatsu *et al.*, 2016; Alakärppä *et al.*, 2018). It has been shown that DNA methylation patterns vary between individuals of a maize population (Eichten *et al.*, 2013). To analyze whether variations contribute to heterosis formation in a *Brassica napus* population, DNA methylation profiles on single-base resolution (RRBS) were generated of a subset (26 hybrids and their respective parents) of a *Brassica napus* breeding population. The DNA methylation patterns of the paternal, maternal and hybrid lines summarized in groups were characterized. The mean cytosine methylation rates were slightly lower than DNA

methylation rates revealed by genome-wide methylation profiling of *Brassica napus* inbred lines (ZY50 and MB) and their hybrid (ZYZ108, Shen *et al.*, 2017). These lines showed mean methylation rates of 58.2–61.1 % (CpG), 25.7–27.4 % (CHG) and 10.4–12.6% (CHH) in young flower buds (Shen *et al.*, 2017). The slightly higher mean DNA methylation rate compared to the mean methylation rate analyzed in this work may be due to the analyzed tissues: In *Arabidopsis*, higher DNA methylation rates were observed in inflorescences (Calarco *et al.*, 2012) compared to root and shoot tissues (Widman *et al.*, 2014) as well as whole plants (Cokus *et al.*, 2008).

The methylation rates of the cytosine contexts were strongly correlated with each other ($r > 0.9$) reflecting the results of Chen *et al.* (2015) in *Brassica rapa* suggesting a tight link of maintenance mechanisms and common RdDM. In the hybrid group, a trend of lower mean methylation rates was found, although these values should be treated with caution as there was a moderate negative correlation ($r < -0.5$) visible between mean methylation rate per line and cytosine coverage (appearing like the higher the coverage, the more regions with lower methylation rates were covered). As the coverage of individual samples was evenly distributed over each group, it can be assumed that there should not be a big bias, or the same bias per group, which creates comparability. The methylation rates did not correlate with the trimmed sequence read count or bisulfite conversion rate, but with mapping efficiency, which therefore marks the point in the analysis chain, where a potential relationship of coverage and mean methylation rate could have started, probably due to the mapping of only unique reads that could introduce a bias.

Although not significant, hybrid lines tended to display lower mean methylation rates compared to maternal lines and paternal lines. Moreover, although total mean methylation rates increased, mean methylation rates in coding regions decreased from low to high heterotic hybrids. Together with the results discussed in 4.2.1, where an additional random demethylation in hybrids could enhance heterosis for seedling growth rate in maize, these results indicate that lower methylation rates at least in coding regions could be beneficial for growth rate heterosis in hybrid seedlings. Other studies showed decreased DNA methylation rates in hybrids compared to their parental lines as well, e.g. in CHG methylation (Greaves *et al.*, 2012) or in CpG islands and retroelements (Raza *et al.*, 2017). Opposing results were revealed by He *et al.* (2010); Shen *et al.* (2012) and Shen *et al.* (2017). Decreased levels of 24 nt sRNA expression probably mediating decrease in *de novo* methylation in hybrids in various species (He

et al., 2010; Groszmann *et al.*, 2011; Barber *et al.*, 2012; Li *et al.*, 2012; Shen *et al.*, 2012) indicate reduction in especially CHH methylation, whereas *Brassica napus* hybrids exhibited higher expression of 24 nt sRNAs (Shen *et al.*, 2017) compared to parental lines. Since no general pattern was found, hybridization effects on DNA methylation in the hybrid cannot be broken down to one responsible overall pattern.

The DNA methylation rates of every contexts differ between A subgenome and C subgenome among the groups but also in total, where the A subgenome is less methylated than the C subgenome. The mean methylation rates of the A subgenome are in the same range as results of studies in *Brassica rapa* (Chen *et al.*, 2015; Niederhuth *et al.*, 2016), which assembles the A subgenome of *Brassica napus*. Moreover, mean methylation rates from *Brassica oleracea* (Parkin *et al.*, 2014; Niederhuth *et al.*, 2016) revealed similar ranges as gained from the C subgenomes in this study. The difference between A and C subgenomes is displayed in the literature (summarized in **Table A.9**) confirming the observed results. However, Shen *et al.* (2017) found more C_n homoeologs expressed at higher level than their corresponding A_n homoeologs of *Brassica napus* suggesting less methylation of the C subgenome as a cause of this higher expression. Since the authors did not analyze whether DNA methylation differences between A and C subgenomes exist, reasons for these indications contrasting the results of this work can be ambiguous. However, in *Brassica rapa* and *Brassica oleracea*, one subgenome (in this case defined as genome copy generated through whole genome duplication events followed by reduction in the *Brassica* evolution) revealed increased DNA methylation levels compared to the the other (Parkin *et al.*, 2014; Cheng *et al.*, 2016). Since the C subgenome was found to be less expressed in resynthesized allotetraploid *Brassica napus* (Wu *et al.*, 2018), the higher DNA methylation rate found on the C subgenome is more likely in the analyzed *Brassica napus* population, too.

4.2.2.1 Differential DNA methylation patterns in a population

Hybridization of epigenetically divergent parental inbred lines can lead to an adjustment or even new composition of epigenetic states in the hybrid. This was suggested to contribute to heterosis formation by several studies (reviewed in Groszmann *et al.*, 2013; Fujimoto *et al.*, 2018). Therefore, differentially methylated positions (DMPs) and regions (DMRs) were called between the parents of each hybrid.

First, at least two different ΔmC s were tested for each cytosine context for DMPs and DMRs and the number of either was compared. DMP numbers did not severely differ with increasing ΔmC , especially in CG contexts, and revealed in total lower numbers than DMRs. Comparing DMR numbers with different ΔmC s revealed a decrease of DMR number with increase of ΔmC . Including all sites of a region depleted the definiteness of DNA methylation patterns due to calculation of a mean, which leads to a certain equilibration. 25 % difference threshold was sufficient for CG, CHG and CHH DMRs, whereas DMPs with that threshold could (almost) not be observed for CHH and CHG. Especially CHH DMPs resided mainly within methylated regions; since these were almost exclusively found in the non-aligned, repetitive portions of the genome (Seymour *et al.*, 2014), the positions of DMPs may not be covered sufficiently. Importantly, DMPs and DMRs do not necessarily coincide (Seymour *et al.*, 2014). Single cytosine positions are less likely to have biological effects (Medvedeva *et al.*, 2014), although there is evidence for function when a transcription factor preferentially binds to a motif with a particular methylation state, e.g the estrogen receptor alpha ESR1 (Fürst *et al.*, 2012). Additionally, differentially methylated CHG and CHH sites were rare; therefore the analysis was continued with DMR data.

Regions, which were differentially methylated between either the maternal and paternal parent with at least 25 % ΔmC for each context were defined as DMRs. In every cytosine context, they showed predominantly dominant DNA methylation patterns in the hybrids. Especially in CHH regions, rather TCdM events were observed in hybrids, which means that hybrids are more adapting to the low parent allele. This is in line with the overall DNA methylation rate decrease at TEs observed in the hybrids.

To analyze the effect of DNA methylation patterns on hybrid performance, the hybrids were divided into equally sized groups of low and high performing hybrids for the *in vitro* seedling trait MPH of total root growth rate. Changes of DMR methylation patterns in hybrids and inheritance were examined between hybrids in the low and high trait value group. Strikingly, the number of both TCM and TCdM events together with the number of dominant and overdominant DNA methylation patterns increased with higher performance while the number of additively methylated loci decreased. This could play a role in heterosis formation in general, as some studies observed an importance of non-additive gene expression for hybrid vigor (Fujimoto *et al.*, 2012; Zhu *et al.*, 2016), which may probably be an effect of non-additive DNA methylation patterns of regions.

4.2.2.2 DNA methylation pattern analysis on genome annotations

DNA methylation is an important regulator of transposon and gene expression. To determine the effect of DNA methylation differences between parents and hybrids on features to whose regulation DNA methylation might contribute in heterosis, cytosine methylation rate per context was analyzed at and closeby annotated features (coding sequences (CDS) and transposable elements (TEs)).

The distribution of DNA methylation rates of different contexts on genome annotations per group divided by subgenome was examined. Similar tendencies compared to genome-wide methylation were revealed (4.2.2): the A subgenome was less methylated in every context at both, CDS and TE annotations, than the C subgenome indicating an interdependent dominance of the larger C subgenome that needs to be compensated (Pelé *et al.*, 2017).

Distribution of mean DNA methylation rates (in 100 bins) over CDS including regions 1 kb up- and downstream revealed a typical pattern for genes showing lower methylation rates in the transcribed regions than their flanking regions (Chen *et al.*, 2015; Junaid *et al.*, 2018). DNA methylation present especially in transcription start (TSS) and termination sites (TTS) as well as promoter regions is known to regulate transcription (Jacobsen & Meyerowitz, 1997; Cubas *et al.*, 1999; Manning *et al.*, 2006), therefore the methylated cytosines present in these regions might have a direct effect on expression of respective genes. The general pattern displays a mixture of promoter-methylated genes and the following classes of gene body methylation summarized by Bewick & Schmitz (2017): 1. Unmethylated genes that are highly expressed, 2. Gene body methylated genes, which show enrichment in CG methylation and depletion in TSS and TTS leading to moderate expression, 3. Genes revealing enriched CG methylation also in TSS, which are usually repressed, 4. Genes, that are enriched in CHG but depleted in CHH methylation leading to a lower expression level than all genes, 5. RdDM genes that are enriched in CHH methylation and whose expression is normally silenced.

The biggest proportion of annotated genes covered with cytosine data may be the unmethylated ones, as DNA methylation rates were rather low compared to up- and downstream of the transcribed regions. Gene body methylation, where methylated CGs are enriched in the transcribed region compared to TSS, TTS and up- and downstream regions, modulates transcription and requires mainly CMT3 (Bewick *et al.*, 2016). It is still not completely solved, how gene body methylation acts on transcription activity of genes, as positive (Ball *et al.*, 2009) but also no (Bewick *et al.*, 2016) or

“complicated” (Raza *et al.*, 2017) correlation with the respective gene expression has been put forward. However, certain levels of especially CG methylation present in transcribed regions might repress transcription (Yang *et al.*, 2014). CMT3 was actually lower expressed than the methyltransferases MET1, CMT2, DRM1 and the chromatin remodeler DDM1 in the data fitting the apparent majority of low/unmethylated genes analyzed. But a certain level of CG and CHG methylation was present in transcribed regions, hence an effect of this DNA methylation pattern cannot be excluded (genes belonging to class 2 and 4). Class 3 seems to make up the lowest proportion of genes shown, as no enrichment in TSS regions was visible meaning that this must have been fully compensated by the other classes. No indication was found for CHH enrichment in transcribed regions, which could potentially be a consequence of reduced RdDM pathway activity within the gene body of the transcribed genes compared with TE (Junaid *et al.*, 2018). However, it is hard to tell about the different classes of gene methylation as all together are displayed. Moreover, hybrids displayed an intermediate CDS methylation level indicating that extremes in cytosine methylation rates of transcribed regions are not beneficial for plant growth. This result is supported by the work of Thiemann *et al.* (2014), who examined the influence of additively expressed, linear with heterosis-correlated genes on heterosis for grain yield in maize by a meta-analysis approach comprising different molecular and phenotypic datasets. The authors proposed that additive expression levels in the hybrid are compensating for detrimental expression in one or the other parent. The cumulative effect of overall more balanced gene expression in hybrids could contribute to heterosis (Thiemann *et al.*, 2014). Furthermore, the “bucket principle” of Cheng *et al.* (2016), that was proposed on heterosis and allopolyploid formation, assumed that the amount of water, which can be hold by a wooden barrel, is determined by its shortest plank. They further considered differential gene expression as the planks and the performance of a protein complex or biochemical pathway as the barrel, which is dependent on the components expressed at the lowest level relative to the need or mutated genes with low function efficiency (shortest plank). When expression of different rate-limiting genes, which are considered as the shortest planks in the parents, combines in the hybrid and adjust to mid-parental values, the strong influence of the “shortest plank” is lessened. Therefore, mid-parental gene expression induced by mid-parental gene methylation could contribute to hybrid vigor.

TE methylation rates in contrast were highly elevated compared to the flanking regions. Cytosine methylation distribution in 100 bins over TEs including 1kb up- and

downstream regions at the A and C subgenome revealed a strong increase of DNA methylation rates of all cytosine contexts. DNA methylation is important for TE silencing in plants (Saze *et al.*, 2012) to ensure genome stability by preventing TE expression and hence transposition into other regions. This repression is mediated by a double lock mechanism using a reinforcement loop between histone methylation H3K9me2 and DNA methylation involving 24 nt siRNAs (reviewed in Slotkin & Martienssen, 2007). Gene expression can be directly affected by TE methylation, for example seen for FWA (FLOWERING WAGENINGEN) expression. Lack of cytosine methylation of a TE upstream of the gene leads to expression increase resulting in a late flowering phenotype (Soppe *et al.*, 2000). RdDM influences TE-regulated gene expression, too, as siRNA-targeted TEs are shown to associate with reduced gene expression in both *Arabidopsis thaliana* and *Arabidopsis lyrata* (Hollister *et al.*, 2011). Hybrids revealed a trend to be less methylated in TE regions especially at the A subgenome, which could cause higher transposon activity in the hybrid. Suggesting similar effects as seen at the FWA locus, less TE methylation could enhance the expression of genes nearby the affected TEs. Hypomethylation of DNA in *ddm1*-derived epiRILs of *Arabidopsis* induced TE mobilization that accumulated over the generations (Quadrana *et al.*, 2018). Especially the mobilized *ATCOPIA93* elements preferably were inserted in the euchromatin and, guided by histone variant H2A.Z, biased in affecting the expression of stress response genes. These findings suggest that TE demethylation and mobilization in hybrids could induce phenotypic differences. The growth limiting genes revealing insertions were further silenced via DNA methylation in later generations (Quadrana *et al.*, 2018) providing insights into possible mechanisms of inbreeding depression.

Correlation analysis of DNA methylation features (genome-wide, on CDS or TEs) and *in vitro* root/shoot traits or field traits based on Pearson correlation coefficients revealed a moderate positive correlation of DMR numbers (total, at CDS and TEs) of each context and oil content in mature plants. Therefore, regions differentially methylated between the parents could have an effect on this trait in the hybrid. Thus, the number of DMRs, rather than the methylation differences of the DMR between the parents could have a growth promoting effect in the hybrid. Looking more in detail at directions of differential DNA methylation and hybrid patterns in these regions, no correlations of field traits (especially oil content) with distinct DNA methylation pattern were revealed in the hybrid (neither TCM nor TCdM or any other) assuming that there is a diversity of patterns generated in the hybrid through the parental methylation differences. In contrast, total root and tiproot growth (including growth rate and MPH of growth rate)

showed moderate positive correlations with CG TCdM events for the CG context at TEs and furthermore both, TCM and TCdM events together, on TEs. These data support a relevance of (TE) hypomethylation in hybrids to promote growth traits, which is supported by the results of Sun *et al.* (2015), who found an increased number of TCdM events in maize hybrids by MSAP. Lauss *et al.* (2018) analyzed heterosis of *Arabidopsis* hybrids generated from *ddm1* mutants-derived (Col background) epiRILs (paternal parent) crossed with Col wildtype (maternal parent). The authors observed that TCdM events preferentially occur at parental DMRs (independent from sequence polymorphisms), whereas some TCM events occurred at regions without parental difference in DNA methylation (Lauss *et al.*, 2018), suggesting *de novo* DNA methylation via RdDM at these loci. In conclusion, not all TCM events may be displayed in the data, although the general trend of hypomethylation suggests rather TCdM events at TEs. Additionally, overdominant DNA methylation patterns in hybrids correlate with the same seedling root/shoot traits suggesting that the observed TCdM events in the hybrid not only adapt to a parental state, but are even less methylated than the lower parent.

4.2.2.3 DNA methylation effect on mRNA and sRNA expression

As discussed before (4.2.2.2), the silencing effect of DNA methylation on gene expression has been shown by several studies in various species (Jacobsen & Meyerowitz, 1997; Cubas *et al.*, 1999; Soppe *et al.*, 2000; Manning *et al.*, 2006; He *et al.*, 2010; Yang *et al.*, 2014). This was confirmed by comparing expression levels of transcripts and DNA methylation rates of cytosine contexts from RRBS data for every location covered in both datasets. As no mRNA expression could be observed at high methylated loci and high expression levels were only detected with very low or without cytosine methylation, which was true for every context, a transcriptional repressing effect of DNA methylation could be confirmed.

Parental differentially expressed genes (DEGs) were identified to analyze their effect on hybrid mRNA expression patterns as it was performed for DNA methylation. Mostly mid-parental (additive) mRNA expression levels in hybrids were observed, which is in line with a majority of studies (Guo *et al.*, 2006; Swanson-Wagner *et al.*, 2006; Meyer *et al.*, 2012; Alonso-Peral *et al.*, 2017). Dominant expression patterns in hybrids were observed in less transcripts and overdominant mRNA expression patterns were even less frequently occurring. Non-additive expression is suggested to be established mainly in *trans* (including transcription factors etc.), whereas additive expression can derive from both, *cis* and *trans* regulation (Springer & Stupar, 2007). The number of both pat-

terns was changed from low to high performing hybrids for MPH of total root growth rate, suggesting that both play a role in heterosis of *Brassica napus*. As proposed by Cheng *et al.* in 2016 (4.2.2.2), additive expression patterns in the hybrid might contribute to heterosis referring to the “bucket principle”. But also non-additive expression patterns are associated to heterosis, e.g. in early *Arabidopsis* developmental stages (Zhu *et al.*, 2016; Alonso-Peral *et al.*, 2017). Increased numbers of expressed genes that were observed in hybrids (Li *et al.*, 2018) would correspond to non-additive expression. However, different levels of loci showing each pattern were revealed in different studies, which probably results from the analysis methods used.

Overlaps between DMRs and DEGs were analyzed to examine whether parental differential methylation has a direct effect on parental differential expression. Only few DMRs per context overlapped with DEGs, suggesting no effect, but *trans* effects as regulators of the genes that are differentially expressed between the parents cannot be excluded with this analysis (e.g. enhancer elements, siRNAs or transcriptions factors). In contrast to Li *et al.* (2018), who found an increased number of expressed genes in *Brassica oleracea* hybrids, which was reflected in DEG analysis, *Brassica napus* hybrids showed a smaller number of expressed transcripts in hybrids with in contrast higher expression. In DEGs, the mRNA expression was significantly reduced in hybrids compared to parental lines, suggesting adaption of expression patterns to the parent with lower expression. This is opposing to hybrid DNA methylation patterns, which showed a mean methylation rate decrease in regions of parental differential methylation that would suggest an increase in expression of nearby genes. An interesting question would be, whether DEGs rather lead to down-regulation in the hybrid whereas DMRs are responsible for up-regulation of gene expression. This would explain why just a few are overlapping.

Other important regulators of gene expression are small RNAs. They can act *in trans* e.g. by post-transcriptional gene silencing in RNAi (Castel & Martienssen, 2013) or as small guiding RNAs in the RdDM pathway (Lewsey *et al.*, 2015). The latter is mediating *de novo* DNA methylation at genes and transposons with 24 nt siRNAs. They are generated mainly at repetitive elements of the DNA (Slotkin & Martienssen, 2007). Therefore, sRNA populations were characterized by their size and expression in hybrids compared to parental inbred lines. From 18 nt to 30 nt, all sizes of sRNAs mapping to the *Brassica napus* reference genome were observed, with an enrichment of 24 nt sRNAs, implicating a high importance of RdDM and RNAi in gene expression regulation

pathways. Comparison of sRNA expression patterns between paternal as well as maternal parent and hybrid lines revealed a general higher expression assuming more regulatory activity of sRNAs in hybrids, which is supported by the study of Shen *et al.* (2017), who found similar patterns in *Brassica napus* hybrids. In contrast, studies in *Arabidopsis* (Groszmann *et al.*, 2011; Li *et al.*, 2012; Shen *et al.*, 2012) and rice (He *et al.*, 2010) determined a down-regulation of 24 nt siRNAs in hybrids, which contributes to non-additive gene expression. The question arises whether this is a species dependent effect including yet unassessable consequences of allotetraploidy. On the other hand, minor roles in heterosis are assigned to 24 nt siRNAs by several studies using mutants defective in genes of the siRNA biogenesis or RdDM pathway (Kawanabe *et al.*, 2016; Zhang *et al.*, 2016b).

Especially 24 nt siRNAs serve as a buffer against genome instability and influence gene expression by directing DNA methylation via RdDM (Lewsey *et al.*, 2015), e.g. by mediating TCM events. Patterns of sRNA expression and DNA methylation of the same loci were analyzed to figure out, to which extent DNA methylation influences sRNA expression or *vice versa*. Similarly to mRNA expression, no sRNA expression was observed at loci with high methylation rates, which was more pronounced for CHH contexts, and high sRNA expression was only observed at loci with low DNA methylation rates. This was true for all analyzed loci: genome-wide and both annotation features, CDS and TEs. The pattern observed rather indicates sRNA origin than target sequence.

Correlations of sRNA expression loci covered with cytosine methylation data or DMRs (total, at CDS, at TEs) of the three contexts with field traits of mature plants and *in vitro* root/shoot seedling traits of hybrids were performed to analyze the relation of sRNA-dependent DNA methylation traits and heterosis as well as hybrid performance. Total sRNA expression revealed a slight positive correlation with all root/shoot traits except root branching angle and root-to-shoot ratio, which fits to the mild negative correlation of the same traits with cytosine methylation present at sRNA producing loci and therefore most likely inhibiting their expression. Taken together, the results indicate a relevance of sRNA expression in heterosis (enhancement) and hybrid performance of seedlings roots in *Brassica napus*.

4.2.2.4 Analysis of genes involved in regulatory pathways

Differences in DNA methylation and sRNA expression were detected between hybrids and parents, which could be a result of changes in cytosine methylation and expression

of specific genes involved in methylation pathways and mechanisms probably involved in growth heterosis, e.g. stress response via the salicylic acid (SA) pathway (Groszmann *et al.*, 2015). Maintenance and *de novo* methyltransferases (MET1, CMT3, DRM2), genes involved in RdDM (AGO4), the chromatin remodeler DDM1 and the DNA glycosylase ROS1, which is known to be involved in stress response (Kim *et al.*, 2019) were analyzed for mRNA expression and DNA methylation patterns in parents and hybrids. The transcription factor CCA1, which is involved in the plant circadian clock regulation (reviewed in Green & Tobin, 2002) and shown to differ in hybrids compared parental inbred lines (Ni *et al.*, 2009), and genes active in stress response, e.g. in the SA biosynthesis and signaling pathway (C4H, MOS1, ICS1, NPR1), were selected to represent important pathways, which may be involved in heterosis formation. All gene loci were tested for overlap with cytosine contexts covered in RRBS data, mRNA expression, sRNA expression, DMRs and DEGs. Neither DMRs nor DEGs could be detected in or close by the selected gene loci. This could be either due to insufficient data coverage or to a dispensible relevance of parental differences in the regulation of these genes. However, a trend of down-regulation from parents to hybrids of analyzed methyltransferases was observed indicating less methylation activity, both maintenance and *de novo*. DRM2, which is the responsible methyltransferase in the RdDM pathway, revealed a significant decrease in mRNA expression in the hybrids compared to both parents suggesting that the RdDM pathway may be decreased and especially short TEs in the euchromatin undergo hypomethylation probably pronounced at the CHH contexts, as *drm2* mutant-derived hypomethylation was negatively correlated with TE size at chromosome arms (Zemach *et al.*, 2013). This could increase gene expression, as demethylation of TEs is known to induce activation of adjacent genes (Hollister & Gaut, 2009). Another indication for that regulational change is the significant decrease in mRNA expression (paternal lines to hybrids) of AGO4, which builds a complex with the guiding 24 nt siRNAs during RdDM, therefore RdDM could be reduced. Moreover, the mRNA expression of both DRM2 and AGO4 revealed negative correlation with some root/shoot seedling traits. The TE activating mechanism might be enhanced in the hybrid by the increased activity of ROS1 that is responsible for demethylation of transposons close to genes, too (Schumann *et al.*, 2017). ROS1 expression is induced by DNA methylation (Bartels *et al.*, 2018), more precisely by methylation of a sRNA target site (RdDM target) in its promotor region. This functions antagonistically to a helitron TE in the ROS1 promotor inhibiting expression through its inactive state caused by methylation as shown in *Arabidopsis* (Lei *et al.*, 2015). ROS1

revealed a higher expression in hybrids compared to their parental inbred lines, together with a higher CG and CHG (but lower CHH) methylation indicating increased activity of *de novo* methylation at least at CG and CHG contexts at the ROS1 locus that in turn increases its expression. Targets of ROS1 overlap with siRNAs, which shows the potential reversibility of RdDM (Matzke *et al.*, 2015). As both, sRNA and ROS1 expression were increased in the hybrid, the up-regulation of one could be a result of up-regulation of the other or *vice versa*. ROS1 is also important in plant immunity (López Sánchez *et al.*, 2016), where TEs directly play a role, too (Espinás *et al.*, 2016).

In contrast to DRM2 and AGO4, the chromatin remodeler DDM1 expression was up-regulated in hybrids compared to parental lines. DDM1 is involved in DNA methylation in a RdDM-independent manner and it preferentially targets heterochromatin (Zemach *et al.*, 2013). Moreover, DDM1 is also able to modulate gene expression by regulating adjacent TEs (Lippman *et al.*, 2004). DDM1 works together with CMT2, CMT3 and DRM2, by allowing them to access heterochromatin, as shown in rice (Tan *et al.*, 2018) and *Arabidopsis* (Zemach *et al.*, 2013). DDM1 was already strongly expressed in the paternal lines compared to the other genes, but even slightly increased expression in hybrids. Since neither increased expression of CMT2 or CMT3 nor enhanced genome-wide DNA methylation was observed in hybrids, DDM1 might not only function in increasing general DNA methylation by enhanced expression in hybrids compared to parents. DDM1 also acts antagonistically to RdDM, as shown in rice (Tan *et al.*, 2016). Therefore, increased expression could diminish RdDM to a certain extent, which would be in line with the idea of decrease in TE methylation to increase or activate transcription of neighboring genes. In rice, DDM1 facilitates CHH methylation via DRM2 (Tan *et al.*, 2018), which is not consistent with the results of the *Brassica napus* subset, where DDM1 is increased, but CHH methylation showed a slight decrease at TEs. However, DDM1 plays an important role as a genome-wide regulator being involved not only in heterosis-related pathways (Zhang *et al.*, 2016a), but also stress response. Stress-induced expression of DDM1 was shown to reset stress-induced epigenetic states in *Arabidopsis* seedlings to prevent mitotic propagation of these states (Iwasaki & Paszkowski, 2014). Plant stress response and growth heterosis are proven to be linked (Groszmann *et al.*, 2015). This connection involves DDM1, as *ddm1* mutants showed impaired heterosis, combined with increased non-additive expression of genes related to SA metabolism (Zhang *et al.*, 2016a). Therefore DDM1 may function as the epigenetic link between heterosis and SA, a regulator of plant defense and known to play a role in heterosis (Groszmann *et al.*, 2015). Thus, stress-induced hormesis might underly

specific heterosis traits (Zhang *et al.*, 2016a). On the other hand, hypomethylated and demethylated *Arabidopsis* mutants, respectively, showed increase in defense responses, and enhanced RdDM induced in ROS1-defective plants led to decreased resistance (reviewed in Espinas *et al.*, 2016). Hence, ROS1 and DDM1 may target different genomic regions or genes, which may result on one hand in silencing of stress response genes in the hybrid induced by enhanced DDM1 expression. At the same time, increased ROS1 expression, but also decreased DRM2 expression (both of which leading to less RdDM), could induce TE demethylation enhancing adjacent gene expression.

Genes representing SA-related defense response in plants were selected to analyze the impact of aberrant DNA methylation in hybrids on SA metabolism and thus on heterosis formation. SA derives from chorismate through two main pathways: the isochorismate pathway and the phenylalanine pathway (Khan *et al.*, 2015). The genes C4H, which triggers SA production via the phenylalanine pathway, and ICS1, which mediates SA biosynthesis via the isochorismate pathway, exhibited a general trend in down-regulation, in combination with a slight increase in DNA methylation rate, especially at CG and CHG contexts. ICS1 was found to be down-regulated in hybrids of *Brassica napus* (Shen *et al.*, 2012) and *Arabidopsis* (Groszmann *et al.*, 2015) as well. This is further supported by the positive correlation of C4H methylation rate with some root and shoots traits and suggests a regulation of these genes by DNA methylation to contribute to SA metabolism control in hybrids. NPR1, a receptor of the signaling molecule SA as shown in *Arabidopsis* (Wu *et al.*, 2012), activates PR1, the plant defense gene *Pathogen related 1* (Boyle *et al.*, 2009). Therefore, it mediates the SA-related plant systemic acquired resistance upon stress induction. The significant (at least compared to the maternal lines) down-regulation of NPR1 in hybrids confirms a down-regulation of the SA pathway upon hybridization in *Brassica napus*. Plant immunity is highly complex. Sensors of pathogen-derived molecules are nucleotide-binding (NB) domain- and leucine-rich repeat (LRR)-containing proteins (Zhu *et al.*, 2010). SNC1 (suppressor of *npr1-1*, constitutive 1), a Toll-like/interleukin-1 receptor (TIR)-NB-LRR plant *Resistance* protein, is suggested to activate both SA-dependent and SA-independent resistance pathways (Zhang *et al.*, 2003). It is epigenetically regulated by MOS1 (MODIFIER OF SNC1). MOS1 loss was shown to suppress the constitutively active autoimmunity phenotype of the gain-of-function *snc1* mutant via DDM1-induced chromatin modification (Li *et al.*, 2011). The detection of MOS1 and SNC1 transcripts was not possible suggesting either very low levels or poor data coverage, but DNA methylation rates showed increase in CG and especially CHH contexts at the MOS1 locus in hybrids compared to

inbred lines. Unfortunately, SNC1 methylation could not be detected because the data were not covering this locus in the majority of the lines, but an increased CG/CHH methylation rate at the MOS1 locus could indicate an enhanced down-regulation of MOS1 leading to a further decrease of SNC1 expression. This could result in a basal down-regulation of the plant immunity pathways that depend on these genes.

CCA1, *circadian clock associated 1*, which together with *LHY* (*late elongated hypocotyl*) regulates circadian clock oscillation, was shown to be altered in hybrids and polyploids (Ni *et al.*, 2009). *CCA1* expression is epigenetically repressed during the day, which results in induction of downstream genes that contain evening elements in chlorophyll and starch metabolic pathways in hybrids, thus enhancing production of these in hybrids (Ni *et al.*, 2009). Altered expression amplitudes in hybrids correlate with biomass heterosis in *Arabidopsis* (Ng *et al.*, 2014). It is therefore likely playing a role in heterosis formation. *CCA1* revealed an increase in CG methylation rate which might result in a significant decrease (paternal lines to hybrids) in mRNA expression. The DNA methylation rate increase positively correlated with root and shoot seedling traits. This reflects the results of several studies (Ni *et al.*, 2009; Shen *et al.*, 2012; Ng *et al.*, 2014), who proved that *CCA1* expression is not beneficial for plant growth heterosis.

4.2.2.5 Heterosis-associated DMRs

Parental DMRs were associated to MPH of total root growth rate using an approach based on binomial distribution (Seifert *et al.* 2018) to analyze, whether specific parental differentially methylated regions may be involved in the increase of growth heterosis in hybrids. The CG context, 15 DMRs were associated with MPH of total root growth rate, from which two were more often found in high heterotic hybrids (= positive ha-DMR). Seven parental CHG DMRs and one CHH DMR were exclusively negatively associated to MPH of total root growth rate. Parental DMRs, especially in non-CG context, may therefore be rather disadvantageous for heterosis promotion.

As a result, it is possible to assign ha-DMRs with the method of Seifert *et al.* (2018) for every context even from a small population. However, neither methylation nor sRNA expression was significantly differing from low to high heterotic hybrids in these regions. Additionally, the absolute difference of the parental methylation rate within these DMRs was not showing an obvious change in patterning suggesting that the extent of parental difference does not matter. It still needs to be further examined how parental DMRs might play a role in heterosis.

4.3 Factors contributing to Heterosis

How are DNA methylation and mRNA expression (of special genes) organized and, taken together, contribute to growth heterosis formation? Correlations of DNA methylation (and related) marks with high heterotic seedling traits were demonstrated in this work. These correlations were more consistent over the given phenotypic seedling traits compared to field traits, which might be due to factors impacting field traits, e.g. environmental impacts, but also to higher complexity of some field traits, e.g. the antagonistic effect of yield and protein content (Monaghan *et al.*, 2001) that is displayed in the data.

Heterosis is defined as the superior performance of heterozygous offspring compared to their parental homozygous inbred lines. But superiority may depend on the view: human benefit is not necessarily plant benefit. Large scale rewiring equatable with a loss of control caused by hybridization can have various effects, from hybrid incompatibility (Blevins *et al.*, 2017) and hybrid necrosis (Kimatu & Bao, 2016) to increased resistance (Kempf *et al.*, 2017) or growth even in an early developmental stage (Jahnke *et al.*, 2010). Phenotypic variations following hybridization can be induced by epigenetic differences (Dapp *et al.*, 2015; Lauss *et al.*, 2018), although these changes may be due to TE insertions caused by hypomethylation of TEs (Quadrona *et al.*, 2018). The impairment of growth limiting factors (Bar-Zvi *et al.*, 2017), for example down-regulation of stress defense responses in an unstressed state (Miller *et al.*, 2015), and enhancement of growth promoting factors (Bar-Zvi *et al.*, 2017), for example increase of gibberellines and brassinosteroids (Hu *et al.*, 2017), could come together at the same time in the hybrid and lead to growth heterosis formation.

The following chapter summarizes the results and assumptions discussed so far. The hypotheses are summarized in **Figure 4.1** displaying some of the possible mechanisms in hybrids that could lead to the formation of growth heterosis.

One factor that may be involved in heterosis is additive gene expression potentially caused by additive DNA methylation of coding sequences. Several studies showed a majority of genes additively expressed in hybrids (Meyer *et al.*, 2007; Stupar *et al.*, 2008). Although this is not reflected by the mRNA expression data shown here, additive DNA methylation levels were found at annotated coding sequences, which could be a hint, that at least a certain proportion of genes is additively expressed in the hybrids. The rate-limiting effect of some genes could therefore be diminished in the hybrid compared to the parent leading to increased fitness that produces a heterotic pheno-

type. As hybrids are selfed, increasing proportions of their genomes revert to homozygosity like its parents and the respective genes become rate-limiting again (Cheng *et al.*, 2016), which may result in inbreeding depression.

The second factor discussed in this work that might contribute to heterosis is the hypomethylation of transposable elements observed in hybrids. Activation of neighboring genes or repression of genes through insertion could be a cause of TE hypomethylation. Lower methylation rates of cytosine contexts TE regions were observed in hybrids together with decreases of methyltransferase expression of CMT3 that maintains methylation of CHG and CHH contexts and DRM2 that methylates all contexts *de novo* in the RdDM pathway. Additionally, AGO4, another player in RdDM, was down-regulated in hybrids compared to the parental inbred lines. Since no differential expression of CMT2 or MET1 was observed between parents and hybrids, the results indicate a down-regulation of RdDM through hybridization. DRM2-mediated DNA methylation is important in both hetero- and euchromatin, but DRM2 is the only methyltransferase active in euchromatin to silence short TEs (Zemach *et al.*, 2013). Most active genes are found in the euchromatin, as it is the “open”, accessible chromatin state (Fransz & De Jong, 2002; Wang *et al.*, 2006), suggesting that a reduction of DRM2-mediated DNA methylation rather increases expression of already active genes than activates new genes. Although several studies observed down-regulation of siRNAs in hybrids (He *et al.*, 2010; Groszmann *et al.*, 2011; Barber *et al.*, 2012), an increase in sRNA production was observed in *Brassica napus* hybrids, which is in line with other results from *Brassica napus* (Shen *et al.*, 2017), but contradicts with the theory of a reduced RdDM at the first glance. It is not sure, in which mechanism the expressed sRNAs will be functional (e.g. RNAi or miRNAs), as it seems that the loci they aligned to indicate their origin by decreased DNA methylation. A sRNA targeting approach could explain the down-regulation of the methyltransferases in hybrids, since this is not resulting from increased DNA methylation. They are possibly post-transcriptionally silenced, e.g. by RNAi cleavage. On the other hand, the demethylating enzyme ROS1, which was up-regulated in hybrids, is known to require methylation of a sRNA target site in its promoter region for up-regulation (Lei *et al.*, 2015). Taken together, an increased sRNA expression might be useful in the allotetraploid *Brassica napus* to rather silence positive regulators of DNA methylation and increase expression of negative regulators of DNA methylation, resulting in increased gene expression.

A third player in heterosis formation is the chromatin remodeler DDM1. DDM1 was up-regulated in *Brassica napus* hybrids compared to their parental lines. It is central in DNA methylation regulation by facilitating accessibility of chromatin to DRM2 and CMT3 (Higo *et al.*, 2012; Tan *et al.*, 2016, 2018); a lack of DDM1 induces hypomethylation, as shown in several studies (Zemach *et al.*, 2013; Zhang *et al.*, 2016a; Lauss *et al.*, 2018). Because DRM2 expression decreased and cytosine methylation rates were lower in TEs and intermediate in CDS between parents and hybrids, DDM1 increase from parents to hybrids is not fully responsible for transposon silencing in *Brassica napus* hybrids. Instead, in some studies, DDM1 expression was negatively associated to SA-related gene expression, for example of ICS1 (Zhang *et al.*, 2016a; Shen *et al.*, 2017). Furthermore, DDM1 is required for successful interaction of MOS1 with SNC1 (Li *et al.*, 2011) responsible for SA accumulation. Hence, increased DDM1 expression could lead to a basal SA level decrease through different pathways. Additionally, environmentally induced epigenetic changes are mostly transient and not transgenerationally inherited and one of the restorers of prestress epigenetic states is DDM1 (Iwasaki & Paszkowski, 2014). DDM1-mediated DNA methylation is also known to play a role in transposon regulation during meiosis and gametogenesis (Slotkin *et al.*, 2009; Melamed-Bessudo & Levy, 2012). Referred to heterosis, DDM1 may therefore be an important regulator of transgenerational epigenetic resetting involved in the occurrence of inbreeding depression.

A trend becomes visible that especially the basal mechanism of plant stress response is affected in hybrids potentially leading to heterosis. Especially SA-related pathways seem to play a role. SA defends plants from pathogen infection (e.g. by inducing systemic acquired resistance, Vernooij *et al.*, 1994) and various abiotic stresses (reviewed in Khan *et al.*, 2015). Previous findings have suggested that a decreased SA level contributes to the increased growth in hybrid plants (Groszmann *et al.*, 2015; Zhang *et al.*, 2016a; Shen *et al.*, 2017). Furthermore, it was shown that higher SA levels have more phytotoxic effects (War *et al.*, 2011). SA mediates for example the phenylpropanoid pathway: it regulates POD (peroxidase), PPO (polyphenol oxidase), SOD (superoxide dismutase), which are all defense-related, e.g. inducing ROS (reactive oxygen species), which regulate some plant defense genes and are second messengers in hormone signaling pathways (Khan *et al.*, 2015). DNA methylation of the genes C4H, ICS1 and MOS1, which are involved in SA accumulation in separate pathways, could result in a basal decrease in SA metabolism in hybrid plants. Additionally, the SA receptor gene NPR1 was down-regulated revealing the same tendency. The pattern of differed de-

defense-related gene expression observed in F_1 suggests that the hybrid may have decreased capabilities of basal defense response. With regard to the antagonistic relationship between plant immunity and plant growth (reviewed in Wang & Wang, 2014), the lower level of defense-related cellular activity could be a significant factor in allowing for greater growth of the hybrid. For example, the plant growth hormone auxin inhibits SA signaling and in turn, auxin signaling can be inhibited by suppression of an auxin receptor or SA-mediated stabilization of negative regulators of auxin signaling (reviewed in Robert-Seilaniantz *et al.*, 2011). Another example is the suppression of immunity by brassinosteroids, which are growth promoting by signal integration at the level of transcriptional regulation, where brassinosteroid-activated transcription factors negatively regulate immune response (Lozano-Durán *et al.*, 2013). Additionally, the cell proliferation promoting gene DEL1 (DP-E2F-like 1, an E2F transcription factor repressing genes that promote endoreduplication onset) was observed to be influenced by SA level: In *del1-1* mutants, genes involved in SA biosynthesis like ICS1 were elevated leading to an increase in SA production (Chandran *et al.*, 2014). These results illustrate the significance of decreased SA levels for growth enhancement. Additionally, an enhanced growth rate in early developmental stages in *Arabidopsis* mutants lacking glucosinolate (a defense compound of Brassicaceae) was shown by Züst *et al.* (2011). The down-regulation of basal plant stress responses might be part of higher energy efficiency through decrease of protein metabolism (reviewed in Goff, 2011). However, energy use efficiency (EUE) was associated with epigenetic components in hybrids (Hauben *et al.*, 2009): hybrids from isogenic doubled haploid *Brassica napus* lines selected for low respiration rates showed an increase in yield. A general hypomethylation was observed in these low respiration lines, indicating that lower DNA methylation has a positive effect on energy use efficiency.

The question arises why hybrids are not more susceptible to stress, when their basal defense response is down-regulated. It is not clear how molecular mechanisms differing between parents and hybrids change in hybrids from basal to reaction mode during actual stress as data on this seems to be rare. Groszmann *et al.* (2015) demonstrated that a reduction of SA is important for heterosis, but the stress tolerance in the hybrids was not necessarily compromised. Kliebenstein (2016) proposed that plants possibly molecularly adapt directly to the environmental influences which would indicate that the basal state is not of great importance. However, higher endogenous levels of gibberellic acid (GA), which is a growth promoting hormone and is known to affect the stress regulators jasmonic acid (JA), ethylene (ET) or abscisic acid (ABA), were revealed

in hybrids compared to parent lines (Bate *et al.*, 1988; Tsaftaris & Kafka, 2005; Ma *et al.*, 2011). In contrast, maize mutants lacking GA still displayed heterosis (Auger *et al.*, 2005). Therefore, growth heterosis might not depend on enhanced GA levels, thus these could be available in case of needed stress response in the hybrid. Furthermore, the analyzed dataset derived from a *Brassica napus* breeding population, where specific, non-system wide responses, e.g. mono- or polygenic resistances for example to Blackleg (Neik *et al.*, 2017), could be bred leading to similar levels of that special resistances in both, parents and hybrids as was shown for wheat (Thorwarth *et al.*, 2018).

Another factor assumed to play a role in heterosis formation is the circadian clock as differences in DNA methylation and mRNA expression of a gene (CCA1) involved in its regulation appeared between parents and hybrids in *Brassica napus*. The circadian clock optimizes physiology and metabolism in plants and animals (Chen, 2010). In plants, the gene CCA1 (CIRCADIAN CLOCK ASSOCIATED 1), together with LHY (LATE ELONGATED HYPOCOTYL) and TOC1 (TIMING OF CAB 1) regulates the circadian oscillation by a negative feedback loop to produce a self-sustaining and constant periodicity of 24 hours. In *Arabidopsis*, expression of CCA1 increases during the day to decrease to the start level during night (Chen, 2010). CCA1 was shown to suppress starch accumulation and chlorophyll biosynthesis (Ni *et al.*, 2009). Striking changes in DNA methylation and expression of CCA1 between parental lines and hybrids were revealed by Ni *et al.* (2009), Shen *et al.* (2012) and Ng *et al.* (2014) in *Arabidopsis*. The increase in DNA methylation in hybrids leading to a decrease in expression that correlates with increased biomass as observed by Shen *et al.* (2012) was confirmed. Expression regulation was not mediated by parent-of-origin-dependent CHH methylation that (additionally) increased biomass in *Arabidopsis* RdDM mutant hybrids (Ng *et al.*, 2014). In *Brassica napus* F₁, CCA1 down-regulation seems to be rather induced by methylation of CG contexts, as CCA1-associated CHG and CHH contexts were similarly methylated in hybrids and parental lines. However, Ng *et al.* (2014) focussed on changes between reciprocal hybrids, which were not available in the *Brassica napus* population. Subsequently, biomass enhancement through parent-of-origin-dependent CHH methylation rate increase at the CCA1 promotor cannot be excluded. Ng *et al.* (2014) additionally showed that DDM1 has an impact on especially CG methylation at CCA1, what could probably be the driving force of higher CG methylation rates and therefore less expression of CCA1 in the *Brassica napus* hybrids. Finally, their results indicate an

expression change early in embryogenesis of the hybrid, assuming an importance of this developmental stage.

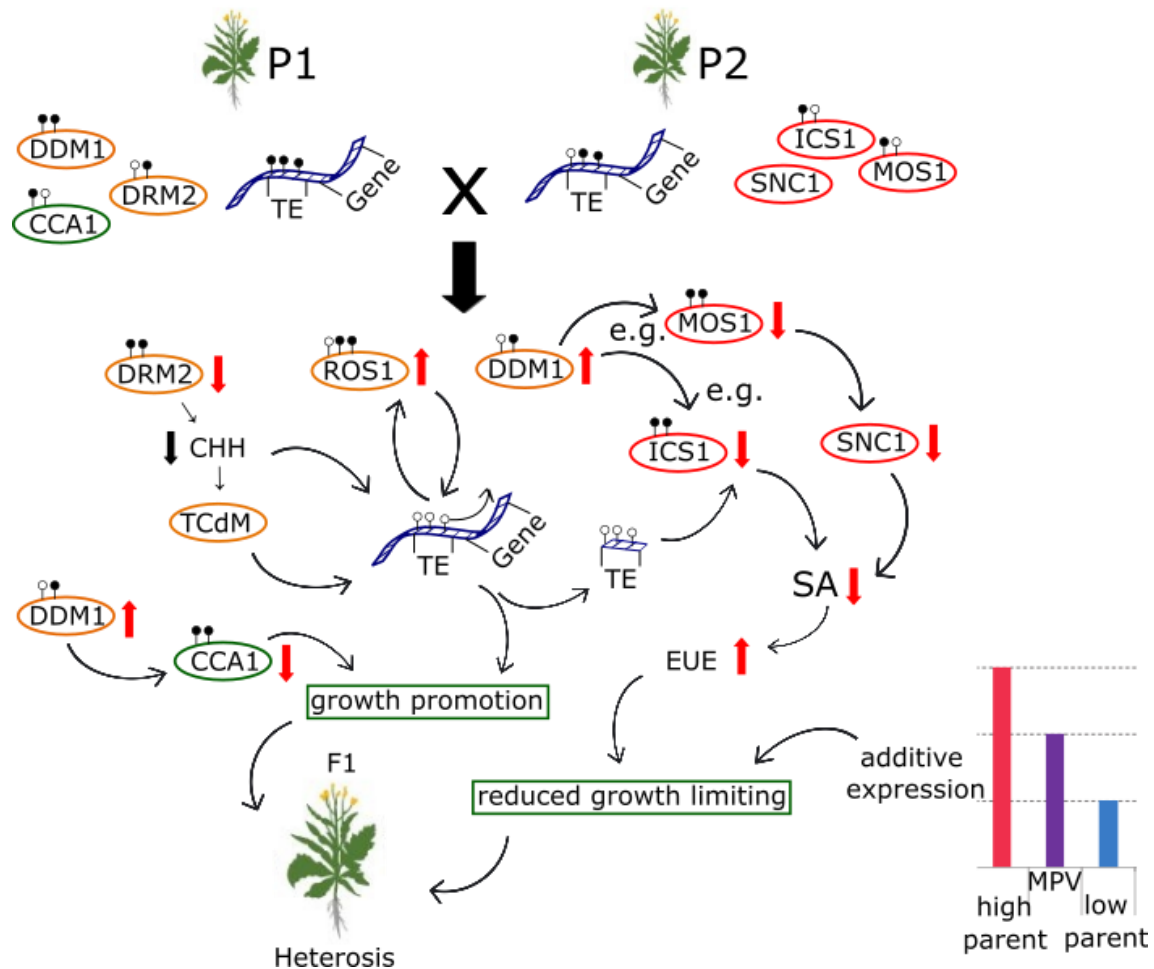


Figure 4.1: Hypothetical model illustrating some of the contributors to heterosis formation.

The additivity illustration element was used from Fujimoto *et al.* (2018).

In conclusion, heterosis is a complex phenomenon, which includes several mechanisms. Some hypothetical players are illustrated in **Figure 4.1**: In parental inbred lines, especially transposons are highly methylated to prevent mobilization and thus genome instability. When two distinct parental (epi)genomes come together through hybridization, large scale rewiring takes place leading to the hypomethylation of (euchromatic short) TEs on one hand and hypermethylation of (SA-related) stress responsive genes and growth-repressing genes (like CCA1) on the other hand. This basal down-regulation of potential growth repressors could increase energy efficiency. The rewiring is mediated by differential expression of methyltransferases, glycosylases and, in the second step, hypomethylation of transposable elements leading to the expression

of adjacent genes on one side and inactivation of genes through transposition activity on the other side. Additive expression induced partly by additive methylation compromises limiting expression levels of important genes. Taken together, enhancement of growth promotion factors and impairment of growth limiting factors may go hand in hand to form growth heterosis, which requires a highly complex interaction of different molecular mechanisms.

Plant tissue and developmental stage differ in regulation. There are hints that transcriptional changes especially early in embryogenesis affects heterosis formation (Ng *et al.*, 2014; Alonso-Peral *et al.*, 2017; 4.2.1). This study provides indications that differential epigenetic patterns especially established early in hybrid development contribute to growth rate vigor. General hypo-/hypermethylation cannot explain heterotic phenotypes; a region- or gene-based regulation is more likely, although this might be a result of genome-wide rewiring. Furthermore, the allotetraploidy of *Brassica napus* might involve changes or even different mechanisms as allopolyploids themselves are observed to display heterotic phenotypes compared to their parental diploid species (Chen, 2010).

There are new opportunities to improve biomass or yield by epigenetic breeding. Hybrid mimics (Wang *et al.*, 2015) were produced by selection for F₁-like phenotype resulting in mostly homozygous breeding lines in the 5th to 6th generation that still exhibit heterosis-like phenotypes. These lines also show down-regulation of stress response genes and differential expression compared to parental lines probably deriving from epigenetic transregulation in the F₁. Recurrent selection for EUE (i.e. low respiration) of *Brassica napus* isogenic lines resulted in enhanced yield, which was stable over five generations (Hauben *et al.*, 2009; De Block & Van Lijsebettens, 2011). EpiRIL/wildtype crossings show heterosis in certain traits (Dapp *et al.*, 2015; Lauss *et al.*, 2018). The findings obtained in this study could be useful for new epigenetic breeding approaches or hybrid breeding, as it identified potential target mechanisms for modification.

5 Perspective

This work confers promising insights into the complexity of mechanisms that work together in the formation of heterosis. However, highly significant changes might not be expected in a genome-wide manner, because calculation of mean values could lead to negligibility of single positions of potential importance. Bisulfite sequencing and qPCR of distinct loci involved in SA-related stress response or TE hypomethylation could validate trends observed through RRBS. Moreover, trends could be possibly confirmed through analysis of a greater population including also reciprocal hybrids.

Another interesting question would be what happens in hybrids under stressed conditions on the molecular level, if the basal stress response is down-regulated, especially the plant stress memory e.g. systemic acquired resistance, which is usually induced by SA.

Maize experiments in this work and studies in *Arabidopsis* (Alonso-Peral *et al.*, 2017) revealed indications that changes in DNA methylation and subsequently gene expression from parents to hybrids during early embryogenesis play a significant role in growth heterosis formation. An experiment comparing molecular patterns of inbred lines and hybrids over different developmental stages starting with the haploid gametes could clarify, when differences occur and further distinguish their impact.

The data sets generated during this project give a variety of further analysis possibilities. Analysis of sRNAs in terms of size and class that are expressed would be helpful for characterizing further players involved in heterosis and find their position in the network. Supportingly, a Gene Ontology analysis could be performed to identify sRNA targets. Classification of hypomethylated TEs (e.g. longterminal repeat retrotransposons or Copia) including density and nearby genes could provide evidence of transposon activation as regulator of growth promoting or growth inhibiting genes. DNA methylation analysis at special loci, especially promoter regions, could further offer information about detailed epigenetic regulatory mechanisms of gene and transposon expression. Moreover, the question whether parental DMRs can be used for prediction of hybrid performance could be tested due to agricultural relevance. Of interest would be, where ha-DMRs are located, what they impact and in which mechanisms they are involved, respectively.

References

- Abdelsamad A, Pecinka A. 2014.** Pollen-Specific Activation of Arabidopsis Retrogenes Is Associated with Global Transcriptional Reprogramming. *The Plant Cell* **26**: 1–16.
- Adams KL, Wendel JF. 2005.** Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology* **8**: 135–141.
- Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. 2012.** MethylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology* **13**: R87.
- Alakärppä E, Salo HM, Valledor L, Cañal MJ, Häggman H, Vuosku J. 2018.** Natural variation of DNA methylation and gene expression may determine local adaptations of Scots pine populations. *Journal of Experimental Botany*.
- Alonso-Peral MM, Trigueros M, Sherman B, Ying H, Taylor JM, Peacock WJ, Dennis ES. 2017.** Patterns of gene expression in developing embryos of Arabidopsis hybrids. *The Plant Journal*.
- Andrews S. 2010.** FastQC A Quality Control tool for High Throughput Sequence Data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Andrews S. 2013.** Reduced Representation Bisulfite-Seq – A Brief Guide to RRBS. *Babraham Bioinformatics*: 1–12.
- Aufsatz W, Mette MF, Van Der Winden J, Matzke AJ., Matzke M. 2002.** RNA-directed DNA methylation in Arabidopsis. *Proceedings of the National Academy of Sciences* **99**: 16499.
- Auger DL, Peters EM, Birchler JA. 2005.** A genetic test of bioactive gibberellins as regulators of heterosis in maize. *Journal of Heredity* **96**: 614–617.
- Axtell MJ. 2013.** Classification and Comparison of Small RNAs from Plants. *Annual Review of Plant Biology*.
- Ball MP, Li JB, Gao Y, Lee JH, Leproust EM, Park IH, Xie B, Daley GQ, Church GM. 2009.** Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature Biotechnology*.
- Bar-Zvi D, Lupo O, Levy AA, Barkai N. 2017.** Hybrid vigor: The best of both parents, or a genomic clash? *Current Opinion in Systems Biology* **6**: 22–27.
- Barber WT, Zhang W, Win H, Varala KK, Dorweiler JE, Hudson ME, Moose SP. 2012.** Repeat associated small RNAs vary among parents and following hybridization in maize. *Proceedings of the National Academy of Sciences* **109**: 10444–10449.
- Bartee L, Malagnac F, Bender J. 2001.** Arabidopsis cmt3 chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Genes and Development* **15**: 1753–1758.

- Bartels A, Han Q, Nair P, Stacey L, Gaynier H, Mosley M, Huang Q, Pearson J, Hsieh T-F, An Y-Q, et al. 2018.** Dynamic DNA Methylation in Plant Growth and Development. *International Journal of Molecular Sciences* 2018, Vol. 19, Page 2144 **19**: 2144.
- Bate NJ, Rood SB, Blake TJ. 1988.** Gibberellins and heterosis in poplar. *Canadian Journal of Botany* **66**: 1148–1152.
- Becker H, Engqvist G, Karlsson B. 1995.** Comparison of rapeseed cultivars and resynthesized lines based on allozyme and RFLP markers. *Theor Appl Genet* **91**: 62–67.
- Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D. 2011.** Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature* **480**: 245–249.
- Benjamini Y., Hochberg Y. 1995.** Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing Author (s): Yoav Benjamini and Yosef Hochberg Source : Journal of the Royal Statistical Society . Series B (Methodological), Vol . 57 , No . 1 Published by : *Journal of the Royal Statistical Society. Series B* **57**: 289–300.
- Bewick AJ, Ji L, Niederhuth CE, Willing E-M, Hofmeister BT, Shi X, Wang L, Lu Z, Rohr NA, Hartwig B, et al. 2016.** On the origin and evolutionary consequences of gene body DNA methylation. *Proceedings of the National Academy of Sciences of the United States of America* **113**.
- Bewick AJ, Schmitz RJ. 2017.** Gene body DNA methylation in plants. *Current Opinion in Plant Biology*.
- Blevins T, Pflieger D, Wang J, Pikaard CS, Pontvianne F. 2017.** Hybrid incompatibility caused by an epiallele. *Proceedings of the National Academy of Sciences*.
- De Block M, Van Lijsebettens M. 2011.** Energy efficiency and energy homeostasis as genetic and epigenetic components of plant performance and crop productivity. *Current Opinion in Plant Biology*.
- Bock C, Tomazou E, Brinkman A. 2010.** Genome-wide mapping of DNA methylation: a quantitative technology comparison. *Nature Biotechnology* **28**: 1106–1114.
- Bologna NG, Voinnet O. 2014.** The Diversity, Biogenesis, and Activities of Endogenous Silencing Small RNAs in Arabidopsis. *Annual Review of Plant Biology* **65**: 473–503.
- Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijó J a, Becker JD. 2008.** Comparative transcriptomics of Arabidopsis sperm cells. *Plant physiology* **148**: 1168–1181.
- Boyle P, Le Su E, Rochon A, Shearer HL, Murmu J, Chu JY, Fobert PR, Despres C. 2009.** The BTB/POZ Domain of the Arabidopsis Disease Resistance Protein NPR1 Interacts with the Repression Domain of TGA2 to Negate Its Function. *The Plant Cell*.
- Calarco JP, Borges F, Donoghue MTA, Van Ex F, Jullien PE, Lopes T, Gardner R, Berger F, Feijó JA, Becker JD, et al. 2012.** Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.**

- 2009.** BLAST+: Architecture and applications. *BMC Bioinformatics*.
- Campenot MK, Zhang G, Cutler AJ, Cass DD. 1992.** Zea mays Embryo Sacs in Culture . I . Plant Regeneration from 1 Day After Pollination Embryos. *American Journal of Botany* **79**: 1368–1373.
- Carré P, Pouzet A. 2014.** *Rapeseed market, worldwide and in Europe*.
- Castel SE, Martienssen RA. 2013.** RNA interference in the nucleus: Roles for small RNAs in transcription, epigenetics and beyond. *Nature Reviews Genetics*.
- Chalhoub B, Denoeud F, Liu S, Parkin I a. P, Tang H, Wang X, Chiquet J, Belcram H, Tong C, Samans B, et al. 2014.** Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. *Science* **345**: 950–953.
- Chan SWL, Henderson IR, Jacobsen SE. 2005.** Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nature reviews Genetics* **6**: 351–360.
- Chandran D, Rickert J, Huang Y, Steinwand MA, Marr SK, Wildermuth MC. 2014.** Atypical E2F transcriptional repressor DEL1 acts at the intersection of plant growth and immunity by controlling the hormone salicylic acid. *Cell Host and Microbe*.
- Chen ZJ. 2010.** Molecular mechanisms of polyploidy and hybrid vigor. *Trends in Plant Science*.
- Chen X, Ge X, Wang J, Tan C, King GJ, Liu K. 2015.** Genome-wide DNA methylation profiling by modified reduced representation bisulfite sequencing in Brassica rapa suggests that epigenetic modifications play a key role in polyploid genome evolution. *Frontiers in plant science* **6**: 836.
- Chen F, He G, He H, Chen W, Zhu X, Liang M, Chen L, Deng XW. 2010.** Expression analysis of miRNAs and highly-expressed small RNAs in two rice subspecies and their reciprocal hybrids. *Journal of Integrative Plant Biology*.
- Cheng F, Sun C, Wu J, Schnable J, Woodhouse MR, Liang J, Cai C, Freeling M, Wang X. 2016.** Epigenetic regulation of subgenome dominance following whole genome triplication in Brassica rapa. *New Phytologist* **211**: 288–299.
- Chettoor AM, Givan S a, Cole R a, Coker CT, Unger-Wallace E, Vejlupkova Z, Vollbrecht E, Fowler JE, Evans M. 2014.** Discovery of novel transcripts and gametophytic functions via RNA-seq analysis of maize gametophytic transcriptomes. *Genome biology* **15**: 1–23.
- Chodavarapu RK, Feng S, Ding B, Simon SA, Lopez D, Jia Y, Wang G-L, Meyers BC, Jacobsen SE, Pellegrini M. 2012.** Transcriptome and methylome interactions in rice hybrids. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 12040–5.
- Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE. 2008.** Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**: 215–219.
- Comai L. 2005.** The advantages and disadvantages of being polyploid. *Nature Reviews Genetics* **6**: 836–846.

- Crow JF. 1948.** *Alternative Hypotheses of Hybrid vigor*. Hanover, New Hampshire .
- Cubas P, Vincent C, Coen E. 1999.** An epigenetic mutation responsible for natural variation in floral symmetry. *Nature*.
- Dapp M, Reinders J, Bédiée A, Balsera C, Bucher E, Theiler G, Granier C, Paszkowski J. 2015.** Heterosis and inbreeding depression of epigenetic Arabidopsis hybrids. *Nature Plants* **1**: 15092.
- Darwin C. 1876.** *The Effects of Cross and Self Fertilisation in the Vegetable Kingdom*. Cambridge: Cambridge University Press.
- Dickinson H, Scholten S. 2013.** And Baby Makes Three: Genomic Imprinting in Plant Embryos. *PLoS Genetics* **9**: 12–14.
- Ding D, Wang Y, Han M, Fu Z, Li W, Liu Z, Hu Y, Tang J. 2012.** MicroRNA transcriptomic analysis of heterosis during maize seed germination. *PLoS ONE*.
- Dong X, Weng Z. 2013.** The correlation between histone modifications and gene expression. *Epigenomics*.
- Dumas C. 1993.** Gametes and Fertilization: Maize as a Model System for Experimental Embryogenesis in Flowering Plants. *THE PLANT CELL ONLINE*.
- East EM. 1936.** Heterosis. *Genetics* **21**: 375–397.
- Edelmann S, Scholten S. 2018.** Bisulfite sequencing using small DNA amounts. In: Bemer M, Baroux C, eds. *Methods in Molecular Biology*. Springer, 45–61.
- Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, Waters AJ, Starr E, West PT, Tiffin P, et al. 2013.** Epigenetic and genetic influences on DNA methylation variation in maize populations. *The Plant cell* **25**: 2783–97.
- Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, Liu S, Yeh CT, Jia Y, Gendler K, Freeling M, et al. 2011.** Heritable epigenetic variation among maize inbreds. *PLoS Genetics* **7**.
- Elhamamsy AR. 2016.** DNA methylation dynamics in plants and mammals: overview of regulation and dysregulation. *Cell Biochemistry and Function*.
- Engel ML, Chaboud A, Dumas C, McCormick S. 2003.** Sperm cells of Zea mays have a complex complement of mRNAs. *Plant Journal* **34**: 697–707.
- Espinas NA, Saze H, Saijo Y. 2016.** Epigenetic Control of Defense Signaling and Priming in Plants. *Frontiers in Plant Science*.
- Feng S, Chen X, Wu S, Chen X. 2015.** Recent Advances in Understanding Plant Heterosis. : 1033–1038.
- Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, et al. 2010.** Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 8689–94.
- Flint-Garcia SA, Buckler ES, Tiffin P, Ersoz E, Springer NM. 2007.** Heterosis is prevalent

- for multiple traits in diverse maize germplasm. *Swedish Dental Journal* **31**: 171–179.
- Fransz PF, De Jong JH. 2002.** Chromatin dynamics in plants. *Current Opinion in Plant Biology*.
- Freeling M, Woodhouse MR, Subramaniam S, Turco G, Lisch D, Schnable JC. 2012.** Fractionation mutagenesis and similar consequences of mechanisms removing dispensable or less-expressed DNA in plants. *Current Opinion in Plant Biology* **15**: 131–139.
- Frisch M, Thiemann A, Fu J, Schrag TA, Scholten S, Melchinger AE. 2010.** Transcriptome-based distance measures for grouping of germplasm and prediction of hybrid performance in maize. *Theoretical and Applied Genetics* **120**: 441–450.
- Frommer M, McDonald LE, Millar DS, Collist CM, Wattt F, Grigg GW, Molloyt PL, Paul CL. 1992.** A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands (genomic sequencing/DNA methylation/bisulfite modification/PCR/kininogen gene).
- Fu D, Xiao M, Hayward A, Fu Y, Liu G, Jiang G, Zhang H. 2014.** Utilization of crop heterosis: A review.
- Fujimoto R, Taylor JM, Shirasawa S, Peacock WJ, Dennis ES. 2012.** Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 7109–14.
- Fujimoto R, Uezono K, Ishikura S, Osabe K, Peacock WJ, Dennis ES. 2018.** Recent research on the mechanism of heterosis is important for crop and vegetable breeding systems. *Breeding Science* **68**: 145–158.
- Fürst RW, Kliem H, Meyer HHD, Ulbrich SE. 2012.** A differentially methylated single CpG-site is correlated with estrogen receptor alpha transcription. *Journal of Steroid Biochemistry and Molecular Biology*.
- Gehring M. 2013.** Genomic imprinting: insights from plants. *Annual review of genetics* **47**: 187–208.
- Goff SA. 2011.** A unifying theory for general multigenic heterosis: energy efficiency, protein metabolism, and implications for molecular breeding. *The New phytologist* **189**: 923–937.
- Grant-Downton R, Hafidh S, Twell D, Dickinson HG. 2009.** Small RNA pathways are present and functional in the angiosperm male gametophyte. *Molecular Plant* **2**: 500–512.
- Greaves IK, Gonzalez-Bayon R, Wang L, Zhu A, Liu P-C, Groszmann M, Peacock WJ, Dennis LS. 2015.** Epigenetic Changes in Hybrids. *Plant physiology* **168**: 1197–1205.
- Greaves IK, Groszmann M, Wang A, Peacock WJ, Dennis ES. 2014.** Inheritance of Trans Chromosomal Methylation patterns from Arabidopsis F1 hybrids. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 2017–22.
- Greaves IK, Groszmann M, Ying H, Taylor JM, Peacock WJ, Dennis ES. 2012.** Trans

chromosomal methylation in *Arabidopsis* hybrids. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 3570–5.

Green RM, Tobin EM. 2002. The role of CCA1 and LHY in the plant circadian clock. *Developmental Cell*.

Griffin PT, Niederhuth CE, Schmitz RJ. 2016. A Comparative Analysis of 5-Azacytidine- and Zebularine-Induced DNA Demethylation. *G3 & Genes/Genomes/Genetics* **6**: 2773–2780.

Groszmann M, Gonzalez-Bayon R, Lyons RL, Greaves IK, Kazan K, Peacock WJ, Dennis ES. 2015. Hormone-regulated defense and stress response networks contribute to heterosis in *Arabidopsis* F1 hybrids. *Proceedings of the National Academy of Sciences*.

Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, Dennis ES. 2011. Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 2617–2622.

Groszmann M, Greaves IK, Fujimoto R, James Peacock W, Dennis ES. 2013. The role of epigenetics in hybrid vigor. *Trends in Genetics* **29**: 684–690.

Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A. 2011. Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature protocols* **6**: 468–481.

Guo M, Rupe MA, Yang X, Crasta O, Zinselmeier C, Smith OS, Bowen B. 2006. Genome-wide transcript analysis of maize hybrids: Allelic additive gene expression and yield heterosis. *Theoretical and Applied Genetics* **113**: 831–845.

Guo W, Zhu P, Pellegrini M, Zhang MQ, Wang X, Ni Z. 2018. CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* **34**: 381–387.

Halldén C, Nilsson N, Rading I, Säll T. 1994. Evaluation of RFLP and RAPD markers in a comparison of *Brassica napus* breeding lines. *Theor Appl Genet.* **88**.

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL, Johnson BE, Fouse SD, Delaney A, Zhao Y, et al. 2010. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nature biotechnology* **28**: 1097–1105.

Hauben M, Haesendonckx B, Standaert E, Kelen K Van Der, Azmi A, Breusegem F Van, Guisez Y, Bots M, Lambert B, Laga B, et al. 2009. Energy use efficiency is characterized by an epigenetic component that can be directed through artificial selection to increase yield. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 20109–14.

He G, He H, Deng XW. 2013. Epigenetic Variations in Plant Hybrids and Their Potential Roles in Heterosis. *Journal of Genetics and Genomics* **40**: 205–210.

He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F, et al. 2010. Global Epigenetic and Transcriptional Trends among Two Rice Subspecies and

Their Reciprocal Hybrids. *The Plant Cell* **22**: 17–33.

Herbst RH, Bar-Zvi D, Reikhav S, Soifer I, Breker M, Jona G, Shimoni E, Schuldiner M, Levy AA, Barkai N. 2017. Heterosis as a consequence of regulatory incompatibility. *BMC Biology* **15**: 1–15.

Higo H, Tahir M, Takashima K, Miura A, Watanabe K, Tagiri A, Ugaki M, Ishikawa R, Eiguchi M, Kurata N, et al. 2012. DDM1 (Decrease in DNA Methylation) genes in rice (*Oryza sativa*). *Molecular Genetics and Genomics* **287**: 785–792.

Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: A trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Research* **19**: 1419–1428.

Hollister JD, Guo Y-L, Smith LM, Ott F, Weigel D, Gaut BS. 2011. Transposable elements and small RNAs contribute to gene expression divergence between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Proceedings of the National Academy of Sciences*.

Hu S, Wang C, Sanchez DL, Lipka AE, Liu P, Yin Y, Blanco M, Lübberstedt T. 2017. Gibberellins Promote Brassinosteroids Action and Both Increase Heterosis for Plant Height in Maize (*Zea mays* L.). *Frontiers in Plant Science*.

Ihaka R, Gentleman R. 1996. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**: 299–314.

Itabashi E, Osabe K, Fujimoto R, Kakizaki T. 2018. Epigenetic regulation of agronomical traits in Brassicaceae. *Plant Cell Reports*.

Iwasaki M, Paszkowski J. 2014. Identification of genes preventing transgenerational transmission of stress-induced epigenetic states. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 8547–52.

Jacobsen S, Meyerowitz E. 1997. Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*.

Jahnke S, Sarholz B, Thiemann A, Kühr V, Gutiérrez-Marcos JF, Geiger HH, Piepho HP, Scholten S. 2010. Heterosis in early seed development: A comparative study of F1 embryo and endosperm tissues 6 days after fertilization. *Theoretical and Applied Genetics* **120**: 389–400.

Jeddeloh JA, Stokes TL, Richards EJ. 1999. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nature Genetics* **22**: 94–97.

Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F. 2012. DNA methylation dynamics during sexual reproduction in *Arabidopsis thaliana*. *Current Biology* **22**: 1825–1830.

Junaid A, Kumar H, Rao AR, Patil AN, Singh NK, Gaikwad K. 2018. Unravelling the epigenomic interactions between parental inbreds resulting in an altered hybrid methylome in pigeonpea. *DNA research : an international journal for rapid publication of reports on genes and genomes* **25**: 361–373.

Kawakatsu T, Huang S, Shan C, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon

- R, Nery JR, Barragan C, He Y, et al. 2016.** Epigenomic Diversity in a Global Collection of *Arabidopsis thaliana* Accessions. *Cell* **166**.
- Kawamura K, Kawanabe T, Shimizu M, Nagano AJ, Saeki N, Okazaki K, Kaji M, Dennis ES, Osabe K, Fujimoto R. 2016.** Genetic distance of inbred lines of Chinese cabbage and its relationship to heterosis. *Plant Gene*.
- Kawanabe T, Ishikura S, Miyaji N, Sasaki T, Wu LM, Itabashi E, Takada S, Shimizu M, Takasaki-Yasuda T, Osabe K, et al. 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**: E6704–E6711.
- Kempf C, Lengeler K, Wendland J, Wendland J. 2017.** Differential stress response of *Saccharomyces* hybrids revealed by monitoring Hsp104 aggregation and disaggregation. *Microbiological Research*.
- Kersey PJ, Allen JE, Allot A, Barba M, Boddu S, Bolt BJ, Carvalho-Silva D, Christensen M, Davis P, Grabmueller C, et al. 2017.** Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Research*.
- Khan MIR, Fatma M, Per TS, Anjum NA, Khan NA. 2015.** Salicylic acid-induced abiotic stress tolerance and underlying mechanisms in plants. *Frontiers in Plant Science*.
- Kim D, Langmead B, Salzberg SL. 2015.** HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**: 357.
- Kim J-S, Lim JY, Shin H, Kim B-G, Yoo S-D, Kim WT, Huh JH. 2019.** ROS1-dependent DNA demethylation is required for ABA-inducible NIC3 expression. *Plant Physiology*.
- Kimatu J, Bao L. 2016.** Epigenetic polymorphisms could contribute to the genomic conflicts and gene flow barriers resulting to plant hybrid necrosis. *African Journal of Biotechnology*.
- Kint S, De Spiegelaere W, De Kesel J, Vandekerckhove L, Van Criekinge W. 2018.** Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. *PLOS ONE*.
- Kliebenstein DJ. 2016.** False idolatry of the mythical growth versus immunity tradeoff in molecular systems plant pathology. *Physiological and Molecular Plant Pathology*.
- Krieger U, Lippman ZB, Zamir D. 2010.** The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature Genetics* **42**: 459–463.
- Krueger F. 2015.** Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
- Krueger F, Andrews SR. 2011.** Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**: 1571–1572.
- Krueger F, Kreck B, Franke A, Andrews SR. 2012.** DNA methylome analysis using short bisulfite sequencing data. *Nature Methods* **9**: 145.
- Langmead B, Salzberg SL. 2012.** Fast gapped-read alignment with Bowtie 2. *Nature*

Methods **9**: 357.

Lauss K, Wardenaar R, Oka R, van Hulten MHA, Guryev V, Keurentjes JJB, Stam M, Johannes F. 2018. Parental DNA methylation states are associated with heterosis in epigenetic hybrids. *Plant Physiology*: 1627–1645.

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

Lee YK, Jin S, Duan S, Lim YC, Ng DP, Lin XM, Yeo GsGS, Ding C. 2014. Improved reduced representation bisulfite sequencing for epigenomic profiling of clinical samples. *Biological procedures online* **16**: 1.

Lei M, Zhang H, Julian R, Tang K, Xie S, Zhu J-K. 2015. Regulatory link between DNA methylation and active demethylation in Arabidopsis. *Proceedings of the National Academy of Sciences*.

Lewsey MG, Hardcastle TJ, Melnyk CW, Molnar A, Valli A, Urich MA, Nery JR, Baulcombe DC, Ecker JR, Jacobsen SE. 2015. Mobile small RNAs regulate genome-wide DNA methylation. *Pnas*: 10.

Li Y, Dong OX, Johnson K, Zhang Y. 2011. MOS1 epigenetically regulates the expression of plant Resistance gene SNC1. *Plant Signal Behav* **6**: 434–436.

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

Li Y, Varala K, Moose SP, Hudson ME. 2012. The Inheritance Pattern of 24 nt siRNA Clusters in Arabidopsis Hybrids Is Influenced by Proximity to Transposable Elements. *PLoS ONE*.

Li H, Yuan J, Wu M, Han Z, Li L, Jiang H, Jia Y, Han X, Liu M, Sun D, et al. 2018. Transcriptome and DNA methylome reveal insights into yield heterosis in the curds of broccoli (*Brassica oleracea* L var . *italica*). : 1–17.

Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**: 471–476.

Liu TJ, Sun LF, Shan XH, Wu Y, Su SZ, Li SP, Liu HK, Han JY, Yuan YP. 2014. Analysis of DNA methylation patterns and levels in maize hybrids and their parents. *Genetics and Molecular Research* **13**: 8458–8468.

López Sánchez A, Stassen JHM, Furci L, Smith LM, Ton J. 2016. The role of DNA (de)methylation in immune responsiveness of Arabidopsis. *Plant Journal*.

Lozano-Durán R, Macho AP, Boutrot F, Segonzac C, Somssich IE, Zipfel C. 2013. The transcriptional regulator BZR1 mediates trade-off between plant innate immunity and growth. *eLife*.

Lu X, Wang W, Ren W, Chai Z, Guo W, Chen R, Wang L, Zhao J, Lang Z, Fan Y, et al. 2015. Genome-wide epigenetic regulation of gene transcription in maize seeds. *PLoS ONE* **10**: 1–20.

- Ma Q, Hedden P, Zhang Q. 2011.** Heterosis in Rice Seedlings: Its Relationship to Gibberellin Content and Expression of Gibberellin Metabolism and Signaling Genes. *PLANT PHYSIOLOGY*.
- Mager S. 2018.** DISSERTATION NUTRITIONAL REGULATION OF DNA METHYLATION AND GENE EXPRESSION IN MAIZE.
- Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, Giovannoni JJ, Seymour GB. 2006.** A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nature Genetics*.
- Martin M. 2011.** Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10.
- Matzke MA, Kanno T, Matzke AJM. 2015.** RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. *Annual Review of Plant Biology*.
- Matzke MA, Mosher RA. 2014.** RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nature Reviews Genetics* **15**: 394.
- Maxwell PH, Burhans WC, Curcio MJ. 2011.** Retrotransposition is associated with genome instability during chronological aging. **108**: 20376–20381.
- Medvedeva YA, Khamis AM, Kulakovskiy I V., Ba-Alawi W, Bhuyan MSI, Kawaji H, Lassmann T, Harbers M, Forrest ARR, Bajic VB. 2014.** Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics*.
- Melamed-Bessudo C, Levy AA. 2012.** Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. *Proceedings of the National Academy of Sciences* **109**: E981–E988.
- Meyer S. 2007.** Heterosis assoziierte Genexpression in frühen Entwicklungsstadien von Zea mays L.
- Meyer S, Pospisil H, Scholten S. 2007.** Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. *Plant Molecular Biology* **63**: 381–391.
- Meyer RC, Witucka-Wall H, Becher M, Blacha A, Boudichevskaia A, Dörmann P, Fiehn O, Friedel S, Von Korff M, Lisek J, et al. 2012.** Heterosis manifestation during early Arabidopsis seedling development is characterized by intermediate gene expression and enhanced metabolic activity in the hybrids. *Plant Journal* **71**: 669–683.
- Miller M, Song Q, Shi X, Juenger TE, Chen ZJ. 2015.** Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of Arabidopsis. *Nature Communications* **6**: 7453.
- Miura F, Enomoto Y, Dairiki R, Ito T. 2012.** Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Research*.
- Moazed D. 2009.** Small RNAs in transcriptional gene silencing and genome defence. *Nature*.

- Moghaddam AMB, Roudier F, Seifert M, Bérard C, Magniette MLM, Ashtiyani RK, Houben A, Colot V, Mette MF. 2011.** Additive inheritance of histone modifications in *Arabidopsis thaliana* intra-specific hybrids. *Plant Journal*.
- Moll R, Lindsey M, Robinson H. 1964.** Estimates of genetic variances and level of dominance in maize. *Genetics* **49**: 411–423.
- Monaghan JM, Snape JW, Chojecki AJS, Kettlewell PS. 2001.** The use of grain protein deviation for identifying wheat cultivars with high grain protein concentration and yield. *Euphytica*.
- Murashige T, Skoog F. 1962.** A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* **15**: 473–497.
- Nagaharu U. 1935.** Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. *Japanese Journal of Botany* **Vol.7** p.3: pp.389-452.
- Neik TX, Barbetti MJ, Batley J. 2017.** Current Status and Challenges in Identifying Disease Resistance Genes in Brassica napus. *Frontiers in Plant Science*.
- Ng DW-K, Miller M, Yu HH, Huang T-Y, Kim E-D, Lu J, Xie Q, McClung CR, Chen ZJ. 2014.** A Role for CHH Methylation in the Parent-of-Origin Effect on Altered Circadian Rhythms and Biomass Heterosis in *Arabidopsis* Intraspecific Hybrids. *The Plant Cell*.
- Ni Z, Kim ED, Ha M, Lackey E, Liu J, Zhang Y, Sun Q, Chen ZJ. 2009.** Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature*.
- Niederhuth CE, Bewick AJ, Ji L, Alabady M, Kim K Do, Page JT, Li Q, Rohr NA, Rambani A, Burke JM, et al. 2016.** Widespread natural variation of DNA methylation within angiosperms. *bioRxiv*.
- Palii SS, Van Emburgh BO, Sankpal UT, Brown KD, Robertson KD. 2008.** DNA Methylation Inhibitor 5-Aza-2'-Deoxycytidine Induces Reversible Genome-Wide DNA Damage That Is Distinctly Influenced by DNA Methyltransferases 1 and 3B. *Molecular and Cellular Biology* **28**: 752–771.
- Pallotta MA, Graham RD, Langridge P, Sparrow DHB, Barker SJ. 2000.** RFLP mapping of manganese efficiency in barley. *Theoretical and Applied Genetics* **101**: 1100–1108.
- Park K, Kim MY, Vickers M, Park J-S, Hyun Y, Okamoto T, Zilberman D, Fischer RL, Feng X, Choi Y, et al. 2016.** DNA demethylation is initiated in the central cells of *Arabidopsis* and rice. *Proceedings of the National Academy of Sciences*.
- Parkin IA, Koh C, Tang H, Robinson SJ, Kagale S, Clarke WE, Town CD, Nixon J, Krishnakumar V, Bidwell SL, et al. 2014.** Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biol* **15**: R77.
- Pearson. 1895.** Notes on Regression and Inheritance in the Case of Two Parents. *Proceedings of the Royal Society of London* **58**: 240–242.
- Pelé A, Trotoux G, Eber F, Lod é M, Gilet M, Deniot G, Falentin C, Nègre S, Morice J, Rousseau-Gueutin M, et al. 2017.** The poor lonesome A subgenome of Brassica napus

var. Darmor (AACC) may not survive without its mate. *New Phytologist* **213**: 1886–1897.

Powers L. 1944. An Expansion of Jones's Theory for the Explanation of Heterosis. *The American Naturalist* **78**: 275–280.

Purty RS, Kumar G, Singla-Pareek S, Pareek A. 2008. *Towards salinity tolerance in Brassica: An overview.*

Quadrana L, Etcheverry M, Gilly A, Caillieux E, Madoui M-A, Guy J, Silveira AB, Engelen S, Baillet V, Wincker P, et al. 2018. Transposon accumulation lines uncover histone H2A.Z-driven integration bias towards environmentally responsive genes. *bioRxiv*.

Quail M a, Otto TD, Gu Y, Harris SR, Skelly TF, McQuillan J a, Swerdlow HP, Oyola SO. 2011. Optimal enzymes for amplifying sequencing libraries. *Nature Methods* **9**: 10–11.

Ranum P, Peña-Rosas JP, Garcia-Casal MN. 2014. Global maize production, utilization, and consumption. *Annals of the New York Academy of Sciences*.

Raza MA, Yu N, Wang D, Cao L, Gan S, Chen L. 2017. Differential DNA methylation and gene expression in reciprocal hybrids between *Solanum lycopersicum* and *S. pimpinellifolium*. *DNA Research*.

Regulski M, Lu Z, Kendall J, Donoghue MTA, Reinders J, Llaca V, Deschamps S, Smith A, Levy D, McCombie WR, et al. 2013. The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. *Genome Research* **23**: 1651–1662.

Riaz A, Li G, Quresh Z, Swati MS, Quiros CF. 2001. Genetic diversity of oilseed *Brassica napus* inbred lines based on sequence-related amplified polymorphism and its relation to hybrid performance. *Plant Breeding* **120**: 411–415.

Robert-Seilaniantz A, Grant M, Jones JDG. 2011. Hormone Crosstalk in Plant Disease and Defense: More Than Just JASMONATE-SALICYLATE Antagonism. *Annual Review of Phytopathology*.

Salvo S, Cook J, Carlson AR, Hirsch CN, Kaeppler SM, Kaeppler HF. 2018. Genetic Fine-Mapping of a Quantitative Trait Locus (QTL) Associated with Embryogenic Tissue Culture Response and Plant Regeneration Ability in Maize (*L.*). *The Plant Genome*.

Saze H, Tsugane K, Kanno T, Nishimura T. 2012. DNA methylation in plants: Relationship to small rnas and histone modifications, and functions in transposon inactivation. *Plant and Cell Physiology*.

Schmidt M, Van Bel M, Woloszynska M, Slabbinck B, Martens C, De Block M, Coppens F, Van Lijsebettens M. 2017. Plant-RRBS, a bisulfite and next-generation sequencing-based methylome profiling method enriching for coverage of cytosine positions. *BMC Plant Biology* **17**: 1–14.

Schnable JC. 2015. Genome Evolution in Maize: From Genomes Back to Genes. *Annual Review of Plant Biology* **66**: 329–343.

Schnable PS, Springer NM. 2013. Progress Toward Understanding Heterosis in Crop

Plants. *Annual Review of Plant Biology* **64**: 71–88.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2012. *The B73 Maize Genome: Complexity, Diversity, and Dynamics* Downloaded from.

Scholten S. 2010. Genomic imprinting in plant embryos. *Epigenetics* **5**: 455–459.

Schumann U, Lee J, Kazan K, Ayliffe M, Wang M-B. 2017. DNA-Demethylase Regulated Genes Show Methylation-Independent Spatiotemporal Expression Patterns. *Frontiers in Plant Science*.

Seifert F, Thiemann A, Grant-Downton R, Edelmann S, Rybka D, Schrag TA, Frisch M, Dickinson HG, Melchinger AE, Scholten S. 2018. Parental expression variation of small RNAs is negatively correlated with grain yield heterosis in a maize breeding population. *Frontiers in Plant Science* **9**.

Seymour DK, Chae E, Grimm DG, Martín Pizarro C, Habring-Müller A, Vasseur F, Rakitsch B, Borgwardt KM, Koenig D, Weigel D. 2016. Genetic architecture of nonadditive inheritance in *Arabidopsis thaliana* hybrids. *Proceedings of the National Academy of Sciences* **113**: E7317–E7326.

Seymour DK, Koenig D, Hagmann J, Becker C, Weigel D. 2014. Evolution of DNA Methylation Patterns in the Brassicaceae is Driven by Differences in Genome Organization. *PLoS Genetics* **10**.

Shen H, He H, Li J, Chen W, Wang X, Guo L, Peng Z, He G, Zhong S, Qi Y, et al. 2012. Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *The Plant cell* **24**: 875–92.

Shen Y, Sun S, Hua S, Shen E, Ye CY, Cai D, Timko MP, Zhu QH, Fan L. 2017. Analysis of transcriptional and epigenetic changes in hybrid vigor of allopolyploid *Brassica napus* uncovers key roles for small RNAs. *Plant Journal*.

Shengwu H, Ovesná J, Kučera L, Kučera V, Vyvadilová M. 2003. Evaluation of genetic diversity of *Brassica napus* germplasm from China and Europe assessed by RAPD markers. *Plant, Soil and Environment*.

Shiferaw B, Prasanna BM, Hellin J, Bänziger M. 2011. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security*.

Shull G. 1908. The composition of a field of Maize. **301**: 12–15.

Shull G. 1948. What is 'heterosis'? *Genetics* **33**: 439.

Siddiq EA, Govila OP, Banga SS. 2004. Heterosis in Crop Improvement. In: Jain HK, Kharkwal MC, eds. *Plant Breeding: Mendelian to Molecular Approaches*. Dordrecht: Springer Netherlands, 451–472.

Sigma-Aldrich. *Imprint DNA modification Kit Manual*.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*.

- Slotkin RK, Vaughn M, Borges F, Tanurdžić M, Becker JD, Feijó JA, Martienssen RA. 2009. Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. *Cell* **136**: 461–472.
- Smallwood S, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J, Andrews SR, Stegle O, Reik W, Kelsey G. 2014. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods* **11**: 817–20.
- Soppe WJJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJM. 2000. The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Molecular Cell* **6**: 791–802.
- Springer NM, Stupar RM. 2007. Allelic variation and heterosis in maize: How do two halves make more than a whole? *Genome Research* **17**: 264–275.
- Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE. 2014a. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nature structural & molecular biology* **21**: 64–72.
- Stroud H, Do T, Du J, Zhong X, Feng S, Patel DJ, Jacobsen SE. 2014b. The roles of non-CG methylation in Arabidopsis. *Nat Struct Mol Biol* **21**: 64–72.
- Stupar RM, Gardiner JM, Oldre AG, Haun WJ, Chandler VL, Springer NM. 2008. Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. *BMC plant biology* **8**: 33.
- Sun LF, Liu TJ, Shan XH, Su SZ, Li SP, Yuan YP, Zhang J. 2015. Analysis of DNA cytosine methylation patterns in maize hybrids and their parents. *Biologia Plantarum*.
- Swanson-Wagner RA, Jia Y, DeCook R, Borsuk LA, Nettleton D, Schnable PS. 2006. All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. *Proceedings of the National Academy of Sciences* **103**: 6805–6810.
- Tan F, Lu Y, Jiang W, Wu T, Zhang R, Zhao Y, Zhou D-X. 2018. DDM1 Represses Noncoding RNA Expression and RNA-Directed DNA Methylation in Heterochromatin. *Plant Physiology* **177**: 1187–1197.
- Tan F, Zhou C, Zhou Q, Zhou S, Yang W, Zhao Y, Li G, Zhou D-X. 2016. Analysis of Chromatin Regulators Reveals Specific Features of Rice DNA Methylation Pathways. *Plant Physiology*.
- Teixeira FK, Colot V. 2010. Repeat elements and the Arabidopsis DNA methylation landscape. *Heredity* **105**: 14–23.
- Thiemann A, Fu J, Seifert F, Grant-Downton RT, Schrag TA, Pospisil H, Frisch M, Melchinger AE, Scholten S. 2014. Genome-wide meta-analysis of maize heterosis reveals the potential role of additive gene expression at pericentromeric loci. *BMC Plant Biology* **14**: 1–14.
- Thorwarth P, Piepho HP, Zhao Y, Ebmeyer E, Schacht J, Schachschneider R, Kazman E, Reif JC, Würschum T, Longin CFH. 2018. Higher grain yield and higher grain protein deviation underline the potential of hybrid wheat for a sustainable agriculture. *Plant*

Breeding.

Tonosaki K, Osabe K, Kawanabe T, Fujimoto R. 2016. The importance of reproductive barriers and the effect of allopolyploidization on crop breeding. *Breeding Science*.

Tsaftaris AS, Kafka M. 2005. Mechanisms of Heterosis in Crop Plants. *Journal of Crop Production*.

Vergeer P, Wagemaker N, Ouborg NJ. 2012. Evidence for an epigenetic role in inbreeding depression. *Biology Letters* **8**: 798–801.

Vernooij B, Uknes S, Ward E, Ryals J. 1994. Salicylic acid as a signal molecule in plant-pathogen interactions. *Current Opinion in Cell Biology*.

Wang L, Greaves IK, Groszmann M, Wu LM, Dennis ES, Peacock WJ. 2015. Hybrid mimics and hybrid vigor in *Arabidopsis*. *Proceedings of the National Academy of Sciences* **112**: E4959–E4967.

Wang W, Meng B, Ge X, Song S, Yang Y, Yu X, Wang L, Hu S, Liu S, Yu J. 2008. Proteomic profiling of rice embryos from a hybrid rice cultivar and its parental lines. : 4808–4821.

Wang Y, Tang X, Cheng Z, Mueller L, Giovannoni J, Tanksley SD. 2006. Euchromatin and pericentromeric heterochromatin: Comparative composition in the tomato genome. *Genetics*.

Wang W, Wang ZY. 2014. At the intersection of plant growth and immunity. *Cell Host and Microbe*.

Wang J, Xia Y, Li L, Gong D, Yao Y, Luo H, Lu H, Yi N, Wu H, Zhang X, et al. 2013. Double restriction-enzyme digestion improves the coverage and accuracy of genome-wide CpG methylation profiling by reduced representation bisulfite sequencing. *BMC genomics* **14**: 11.

War AR, Paulraj MG, War MY, Ignacimuthu S. 2011. Role of salicylic acid in induction of plant defense system in chickpea (*cicer arietinum* L). *Plant Signaling and Behavior*.

Widman N, Feng S, Jacobsen SE, Pellegrini M. 2014. Epigenetic differences between shoots and roots in *Arabidopsis* reveals tissue-specific regulation. *Epigenetics*.

Wu J, Lin L, Xu M, Chen P, Liu D, Sun Q, Ran L, Wang Y. 2018. Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. *BMC Genomics*.

Wu Y, Zhang D, Chu JY, Boyle P, Wang Y, Brindle ID, De Luca V, Després C. 2012. The *Arabidopsis* NPR1 Protein Is a Receptor for the Plant Defense Hormone Salicylic Acid. *Cell Reports*.

Xin M, Yang R, Li G, Chen H, Laurie J, Ma C, Wang D, Yao Y, Larkins BA, Sun Q, et al. 2013. Dynamic Expression of Imprinted Genes Associates with Maternally Controlled Nutrient Allocation during Maize Endosperm Development. *The Plant Cell* **25**: 3212–3227.

Yang H, Chang F, You C, Cui J, Zhu G, Wang L, Zheng Y, Qi J, Ma H. 2015. Whole-

genome DNA methylation patterns and complex associations with gene structure and expression during flower development in *Arabidopsis*. *Plant Journal*.

Yang X, Han H, DeCarvalho DD, Lay FD, Jones PA, Liang G. 2014. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell*.

Yang M, Wang X, Ren D, Huang H, Xu M, He G, Deng XW. 2017. Genomic architecture of biomass heterosis in *Arabidopsis*. *Proceedings of the National Academy of Sciences*.

Yao H, Dogra Gray A, Auger DL, Birchler JA. 2013. Genomic dosage effects on heterosis in triploid maize. *Proceedings of the National Academy of Sciences* **110**: 2665–2669.

Zamore PD, Haley B. 2005. *Ribo-gnome: The Big World of Small RNAs*.

Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. 2013. The arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*.

Zhang Y, Goritschnig S, Dong X, Li X. 2003. A gain-of-function mutation in a plant disease resistance gene leads to constitutive activation of downstream signal transduction pathways in suppressor of npr1-1, constitutive 1. *The Plant cell*.

Zhang M, Kimatu JN, Xu K, Liu B. 2010. DNA cytosine methylation in plant development. *Journal of Genetics and Genomics*.

Zhang Q, Li Y, Xu T, Srivastava AK, Wang D, Zeng L, Yang L, He L, Zhang H, Zheng Z, et al. 2016a. The chromatin remodeler DDM1 promotes hybrid vigor by regulating salicylic acid metabolism. *Cell Discovery* **2**: 1–12.

Zhang L, Peng Y, Wei X, Dai Y, Yuan D, Lu Y, Pan Y, Zhu Z. 2014. Small RNAs as important regulators for the hybrid vigour of super-hybrid rice. *Journal of Experimental Botany*.

Zhang Q, Wang D, Lang Z, He L, Yang L, Zeng L, Li Y, Zhao C, Huang H, Zhang H, et al. 2016b. Methylation interactions in *Arabidopsis* hybrids require RNA-directed DNA methylation and are influenced by genetic variation. *Proceedings of the National Academy of Sciences of the United States of America* **113**: E4248–56.

Zhao X, Chai Y, Liu B. 2007. Epigenetic inheritance and variation of DNA methylation level and pattern in maize intra-specific hybrids. *Plant Science* **172**: 930–938.

Zhu A, Greaves IK, Dennis ES, Peacock WJ. 2017. Genome-wide analyses of four major histone modifications in *Arabidopsis* hybrids at the germinating seed stage. *BMC Genomics*.

Zhu A, Greaves IK, Liu PC, Wu L, Dennis ES, Peacock WJ. 2016. Early changes of gene activity in developing seedlings of *Arabidopsis* hybrids relative to parents may contribute to hybrid vigour. *Plant Journal*: 597–607.

Zhu Z, Xu F, Cheng YT, Wiermer M, Li X, Zhang Y. 2010. *Arabidopsis* resistance protein SNC1 activates immune responses through association with a transcriptional corepressor. *Proceedings of the National Academy of Sciences*.

Zilberman D, Cao X, Jacobsen SE. 2002. Argonaute4 control of locus-specific siRNA accumulation and DNA and Histone methylation. *Science*.

Ziller MJ, Hansen KD, Meissner A, Aryee MJ. 2015. Coverage recommendations for methylation analysis by whole genome bisulfite sequencing. *Nature methods* **12**: 230–232.

Züst T, Joseph B, Shimizu KK, Kliebenstein DJ, Turnbull LA. 2011. Using knockout mutants to reveal the growth costs of defensive traits. *Proceedings of the Royal Society B: Biological Sciences*.

A Appendix

A.1 Tables

Table A.1: Phenotypic field data of the Core factorial.

Black indexes indicate hybrid lines, red maternal lines and blue paternal lines.

index	Thousand kernel weight	yield	Plant length	Beginning of flowering	Field emergence	Oil content	Glucosinolate content	Protein content in seeds
427	4.68	39.22	162.40	109.45	7.67	49.12	14.80	19.76
428	4.98	42.18	168.40	109.33	7.71	49.18	13.22	19.34
429	4.98	44.48	167.19	109.31	7.90	48.27	15.54	19.85
430	5.19	41.68	166.89	107.68	8.13	48.33	13.67	19.48
431	4.53	38.50	163.02	110.75	7.86	51.15	11.33	19.01
432	4.76	42.31	167.72	112.62	7.90	50.18	11.96	19.61
433	4.52	42.53	164.83	111.24	7.98	51.53	11.17	18.94
434	4.54	38.83	165.68	111.62	7.23	50.56	12.70	19.64
435	4.53	41.06	168.37	111.00	7.33	51.14	11.86	19.45
436	4.57	43.00	168.87	111.56	7.54	51.35	11.86	19.47
437	4.59	40.04	170.00	109.97	8.15	50.74	11.10	19.53
438	4.49	39.50	175.41	110.35	7.85	49.84	13.48	19.17
439	4.75	43.73	165.59	109.31	7.37	49.41	11.67	19.33
440	4.67	41.47	162.79	108.22	7.54	49.12	11.35	19.09
441	4.89	43.20	169.53	111.68	8.28	48.68	12.16	19.66
442	4.91	41.58	168.65	111.32	7.54	49.74	11.75	18.95
443	4.81	42.47	168.02	111.29	8.24	49.75	12.02	19.14
444	4.84	41.16	168.34	110.53	7.96	49.64	12.53	19.20
445	4.76	43.30	169.30	109.87	7.88	49.68	11.35	18.96
446	4.50	36.59	166.45	110.58	7.91	49.14	12.11	19.31
447	5.10	45.34	169.86	110.34	7.83	49.52	12.09	19.32
448	4.91	41.39	169.38	111.43	8.11	49.51	12.65	19.24
449	4.53	40.53	168.54	111.62	7.88	49.46	13.17	19.05
450	5.03	44.14	176.25	111.14	7.82	48.54	12.15	19.52
451	4.95	36.91	167.51	109.24	7.80	48.74	11.15	19.62
452	5.17	39.42	162.23	109.67	8.21	48.75	12.02	19.96
453	5.39	42.20	173.15	110.20	7.89	48.52	12.61	19.55
199	4.90	39.15	154.56	109.32	7.03	47.98	17.11	20.49
200	4.80	38.18	164.86	114.03	7.47	51.15	11.99	20.14
202	4.47	38.64	167.03	116.18	7.42	48.89	14.56	19.84
455	5.06	35.46	159.03	112.59	7.74	49.21	13.51	20.29

456								
457	4.59	34.56	160.51	112.20	7.66	48.22	15.65	20.65
458	5.00	34.14	152.94	112.56	7.28	47.53	11.78	20.80
9	4.71	41.53	158.78	113.08	7.42	51.14	11.38	20.01
11	5.31	44.42	156.29	110.75	7.41	51.26	12.76	18.96
17	5.22	36.36	157.52	113.29	6.67	50.19	14.54	20.10
18	4.67	38.06	163.86	110.04	6.86	50.85	15.09	19.62
25	4.97	37.28	159.79	112.25	7.29	49.59	11.60	19.94
29	4.61	31.06	151.18	112.52	7.24	48.71	14.01	21.21
30	5.46	40.85	171.31	112.21	7.81	48.80	18.10	20.17
126	4.98	39.49	171.51	113.49	7.74	47.41	15.10	20.28
132	5.39	42.15	156.32	110.73	7.61	49.16	11.38	19.40
135	5.51	43.54	171.45	113.43	7.33	47.30	11.42	20.25
136	5.48	42.48	168.02	111.92	7.67	50.26	13.22	19.47
454	4.76	24.28	156.88	113.53	7.78	48.39	16.46	20.98
459	5.68	35.44	160.95	107.05	7.33	48.59	13.32	19.57
460	4.61	34.20	150.93	111.80	7.38	50.83	10.73	19.03
461	5.01	34.63	158.18	111.86	7.97	49.44	11.93	20.03
462	5.58	40.17	154.57	110.97	6.44	46.84	11.33	21.41
463	6.15	40.62	160.04	107.00	7.93	48.72	11.82	21.44
464	5.47	44.15	159.67	112.65	7.32	49.46	14.08	20.24
465	4.36	43.72	157.01	115.39	7.27	48.84	12.72	20.34
466	5.45	38.00	155.55	111.75	7.45	50.14	13.05	20.21
467	5.03	37.63	156.54	113.23	6.34	50.07	11.14	19.89
mean_h ybrids	4.80	41.36	167.94	110.42	7.83	49.62	12.35	19.38
mean_p aternal	5.16	38.57	159.83	111.85	7.35	49.33	13.10	20.12
mean_ matern al	4.80	36.69	159.82	112.81	7.43	48.83	14.10	20.37
T-test mat/hy	0.97	0.00	0.00	0.00	0.00	0.09	0.01	0.00
T-test pat/hy	0.00	0.01	0.00	0.00	0.00	0.38	0.09	0.00
T-test mat/pat	0.07	0.37	1.00	0.32	0.65	0.40	0.29	0.41

Table A.2: Phenotypic *in vitro* seedling root and shoot data of the Core factorial.

Following abbreviations were used: Total root growth rate (cm/4 d) = TRGR, Total root length (cm) = TRL, Taproot length (cm) = PRL, taproot growth rate (cm/4 d) = PRGR, Lateral root length_(cm) = LRL, lateral root growth rate (cm/4 d) = LRGR, Root system depth_(cm) = RSD, Root system width (cm) = RSW, Convex_hull_area (cm²) = CHA, Lateral root branching angle (°) = BA, Number of laterals = NRLR, Shoot_dry_weight (mg) = SWD, Root dry weight (mg), Root-to-shoot ratio = R_S. Black indexes indicate hybrid lines, red maternal lines and blue paternal lines.

in-dex	TRG R	TRL	PRL	PRG R	LRL	LRG R	RSD	RSW	CHA	BA	NRL R	SD W	RD W	R_S
427	14.0	198.6	36.9	2.6	161.7	11.4	32.6	9.4	168.7	74.4	134.4	12.2	4.5	0.4
428	4.6	67.2	16.1	1.1	51.1	3.5	14.9	5.3	66.1	51.4	43.5	5.0	2.7	0.4
429	14.7	207.6	35.5	2.4	170.5	12.2	31.1	10.6	209.5	73.1	117.1	12.1	5.2	0.4
430	12.0	173.1	31.2	2.1	141.7	9.8	28.5	10.6	170.4	70.8	85.5	11.3	4.5	0.4
431	8.3	118.2	27.2	1.9	91.0	6.4	25.5	8.6	120.5	75.8	65.6	10.2	3.4	0.3
432	9.0	130.1	28.1	1.9	101.8	7.1	26.3	7.5	120.2	65.2	73.9	10.0	3.2	0.3
433	16.5	233.9	37.3	2.6	194.0	13.7	31.0	12.7	255.6	73.4	122.8	15.2	5.4	0.4
434	15.0	212.6	36.5	2.4	176.0	12.5	32.2	10.7	217.6	74.1	112.0	11.3	5.0	0.5
435	10.8	157.5	29.3	2.0	128.2	8.8	27.5	8.2	131.0	74.1	96.1	11.5	4.2	0.4
436	12.5	176.2	30.0	2.0	146.0	10.5	27.6	10.1	165.1	75.4	90.7	11.9	4.9	0.4
437	9.8	141.3	30.6	2.1	110.6	7.7	27.6	9.4	137.5	73.8	86.6	9.3	3.5	0.4
438	18.5	260.0	44.0	3.1	216.0	15.4	32.8	12.7	273.4	68.9	148.5	15.4	6.4	0.4
439	15.8	219.6	31.8	2.2	185.6	13.5	28.4	11.4	183.8	73.7	111.4	14.1	6.0	0.4
440	9.4	131.3	29.2	1.9	101.6	7.4	27.6	9.5	145.8	73.0	68.9	8.6	4.8	0.5
441	7.1	100.6	21.4	1.5	78.9	5.6	20.0	7.9	88.0	69.4	48.6	7.1	3.0	0.4
442	9.9	138.7	27.9	1.9	109.4	7.9	25.7	9.0	120.2	71.5	63.4	9.7	3.4	0.4
443	13.4	190.5	30.9	2.1	158.2	11.2	28.9	9.9	154.0	74.4	87.5	11.7	5.2	0.5
444	-	-	-	-	-	-	-	-	-	-	-	-	-	-
445	17.4	247.7	41.6	2.9	205.6	14.5	33.2	12.6	265.0	71.0	120.5	14.4	5.2	0.4
446	13.6	193.5	36.3	2.5	156.5	11.1	31.4	12.2	239.2	73.9	100.9	13.4	4.5	0.3
447	14.2	203.7	36.7	2.6	166.4	11.6	28.2	13.9	275.8	63.3	91.3	13.5	5.5	0.4
448	16.0	225.8	36.9	2.6	187.2	13.4	33.0	12.2	246.0	72.7	97.2	14.8	6.0	0.4
449	11.2	157.0	30.0	2.1	124.4	8.9	27.5	13.0	198.5	75.3	77.7	10.0	4.2	0.4
450	14.0	198.6	35.1	2.4	161.1	11.4	30.5	11.9	229.3	73.3	98.7	13.9	4.6	0.3
451	15.3	216.7	31.6	2.1	181.6	12.9	28.4	13.1	198.	72.5	83.1	12.5	5.3	0.4

									8					
452	11.2	158.9	29.1	2.0	127.9	9.2	26.9	9.9	148. 4	72.6	78.2	9.8	5.3	0.6
453	9.6	140.3	26.4	1.8	113.3	7.8	24.9	9.1	129. 8	74.5	66.4	10.3	3.7	0.4
199	6.0	85.5	25.4	1.8	60.1	4.2	23.8	6.5	85.4	73.4	66.8	5.3	2.1	0.4
200	4.6	65.3	21.1	1.5	44.3	3.1	19.6	4.7	50.9	76.2	52.7	5.4	3.0	0.6
202	6.7	95.6	25.3	1.8	70.2	4.9	23.7	7.2	93.9	60.7	57.3	6.5	2.1	0.4
455	4.5	64.4	21.2	1.5	43.0	3.0	19.9	6.2	70.9	75.3	34.7	5.0	2.0	0.4
456	-	-	-	-	-	-	-	-	-	-	-	-	-	-
457	8.7	124.2	32.0	2.3	92.1	6.5	29.2	8.2	144. 7	75.2	76.6	6.2	4.4	0.8
458	10.0	142.2	25.3	1.8	116.5	8.3	23.1	9.6	127. 9	57.5	63.6	8.5	3.6	0.4
9	5.5	77.3	19.6	1.3	57.6	4.2	17.9	7.5	73.8	74.3	42.9	7.0	2.2	0.3
11	7.2	100.3	25.6	1.8	74.6	5.4	24.6	7.7	107. 2	70.9	51.6	5.3	2.4	0.4
17	6.7	91.7	20.3	1.5	71.4	5.2	18.9	9.5	120. 9	52.1	45.0	9.6	2.8	0.3
18	5.9	85.9	24.4	1.7	61.5	4.2	22.6	6.6	81.9	76.5	70.5	6.5	2.3	0.4
25	9.2	132.8	30.7	2.1	102.1	7.0	28.5	7.3	124. 4	72.9	78.7	9.3	1.9	0.2
29	5.1	74.1	20.5	1.4	53.6	3.7	19.1	8.1	76.9	77.0	41.2	6.7	2.2	0.3
30	8.0	113.2	27.8	1.9	85.3	6.1	25.9	8.8	128. 4	74.6	75.8	5.3	3.2	0.6
126	12.2	172.1	31.6	2.2	140.2	10.0	29.2	9.3	159. 5	74.9	109. 0	8.3	3.0	0.4
132	5.7	82.8	23.1	1.6	59.5	4.1	21.3	5.8	70.6	69.0	48.9	4.8	2.1	0.4
135	5.3	75.8	20.9	1.5	54.8	3.8	19.8	8.6	95.3	74.4	35.9	6.8	2.9	0.5
136	7.3	104.3	24.7	1.7	79.3	5.6	22.7	9.9	116. 4	77.1	57.0	6.9	2.8	0.4
454	6.7	98.3	24.9	1.7	73.4	5.0	22.9	7.4	86.2	79.4	68.4	7.3	2.0	0.3
459	7.1	101.8	28.8	2.0	72.4	5.0	26.8	6.7	93.2	74.5	63.2	7.1	2.4	0.4
460	5.1	72.3	20.1	1.4	52.1	3.6	17.6	6.2	78.8	63.8	44.0	3.8	1.7	0.5
461	6.0	85.8	20.6	1.4	64.8	4.6	19.2	9.2	90.0	73.4	34.7	5.0	2.2	0.5
462	7.4	105.1	29.7	2.1	75.3	5.3	27.2	8.5	126. 4	75.3	60.0	5.9	1.9	0.4
463	9.4	129.6	33.8	2.3	95.5	7.0	30.6	11.3	194. 1	72.7	65.6	6.3	3.5	0.6
464	9.0	129.1	33.4	2.3	95.6	6.7	30.2	10.5	174. 7	74.9	64.8	7.9	3.7	0.5
465	5.3	73.8	22.9	1.6	50.7	3.7	21.2	7.8	86.9	74.8	40.8	5.3	2.0	0.4
466	4.9	68.7	18.3	1.3	50.2	3.6	16.4	8.3	80.7	63.6	48.7	5.2	2.7	0.5
467	6.4	90.1	22.9	1.6	67.2	4.8	21.2	8.2	97.5	76.1	54.4	4.5	2.0	0.4
Me an hyb rid	12.4	176.9	31.8	2.2	144.1	10.2	28.2	10.4	179. 2	71.6	91.2	11.5	4.6	0.4
Me an pat	6.91	98.3	25.0	1.7	73.2	5.2	23.0	8.2	107. 8	72.5	57.2	6.4	2.5	0.4
Me an mat	6.8	96.2	25.0	1.8	71.0	5.0	23.2	7.1	95.6	69.7	58.6	6.2	2.9	0.5

p mat /hy	0.00	0.00	0.01	0.03	0.00	0.00	0.01	0.00	0.00	0.47	0.01	0.00	0.00	0.06
p pat /hy	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.59	0.00	0.00	0.00	0.86
p mat /pat	0.86	0.87	0.97	0.85	0.84	0.81	0.93	0.10	0.45	0.37	0.86	0.72	0.21	0.20

Table A.3: Gene loci of *Brassica napus* genes used for comparison with DNA methylation, mRNA and sRNA data.

Gene loci were obtained from <https://www.ncbi.nlm.nih.gov/gene> (10/12/2018).

Name	abbreviation	chromosome	Start	Stop	Gene ID
DNA (cytosine-5)-methyltransferase CMT3-like	CMT3	chrA06	22826063	22831080	LOC106351898
DNA (cytosine-5)-methyltransferase 1-like	MET1	chrC08	37567598	37573820	LOC106415488
DNA (cytosine-5)-methyltransferase DRM2-like	DRM2	chrC09	46029969	46034101	LOC106418406
ATP-dependent DNA helicase DDM1-like	DDM1	chrA09	5059196	5063383	LOC106366290
Argonaute 4	AGO4	chrA04	14704669	14710600	LOC106446773
ROS1-like	ROS1	chrA05	6293438	6301651	LOC106451111
CCA1-like	CCA1	chrA05	727731	730273	LOC106450471
Trans-cinnamate 4-monooxygenase-like	C4H	chrA06	24210151	24212169	LOC106348398
MODIFIER OF SNC1 1-like	MOS1	chrA01	10044130	10046943	LOC106359205
regulatory protein NPR1	NPR1	chrC01	45101508	45109602	LOC106389246
isochorismate synthase 1	ICS1	chrA10	5790077	5793940	LOC106355795
isochorismate synthase 1	ICS1	chrA07	17138029	17141886	LOC106446194

Table A.4: Analysis of heterosis for seedling growth rate for hybrids (A188 x H99) and inbred lines (A188 and H99) grown *in vitro* from aza-treated and non-treated embryos.

	BPH		MPH
	- aza	+ aza	- aza
	0.50	0.99	0.58
	0.81	1.43	0.94

	1.12	1.37	1.29
	0.70	1.02	0.81
	1.34	1.73	1.55
	0.63	1.31	0.73
	0.50	0.13	0.57
	0.89	1.32	1.03
	1.26	0.88	1.46
	0.48	1.55	0.56
	1.40	1.50	1.61
	1.11	1.41	1.28
	0.82	1.73	0.94
	0.77	1.08	0.89
	0.92	0.93	1.06
	0.99	1.19	1.15
	1.21	1.32	1.40
	0.49	1.54	0.57
	0.74	1.47	0.86
	1.06	1.24	1.22
	0.70	0.56	0.80
	1.49	1.52	1.72
	1.34	1.86	1.55
	1.43		1.66
	1.35		1.56
	0.39		0.45
mean	0.94	1.26	1.0866
standard deviation	0.34	0.39	0.39
standard error of mean	0.29	0.29	0.34

Table A.5: Basic statistics of RRBS data of the *Brassica napus* “Core factorial”. Methylation (CpG, CHG, CHH, in %) and coverage (Ccov CG, Ccov CHG, Ccov CHH, in %) of the three cytosine contexts were calculated with 5 reads/cytosine.

Mapping efficiency is given as ME and conversion rate is given as CR. Averages and standard deviation is calculated for each group at the end of the table. Black indexes indicate hybrid lines, red maternal and blue paternal lines.

Sample	Trimmed reads	CR cp [%]	ME [%]	CpG	CHG	CHH	Ccov CG	Ccov CHG	Ccov CHH	Ccov total
9	10314623	99.28	38.50	46.08	17.15	4.35	3.43	5.00	3.63	3.80
11	9570612	99.31	37.30	49.77	20.70	4.96	2.59	3.67	2.52	2.70
17	14442388	99.52	43.60	40.73	16.55	3.86	6.68	9.50	6.96	7.30
18	15090860	99.54	42.60	40.89	13.86	3.38	7.35	10.52	7.89	8.20

25	10336281	99.32	38.20	45.33	16.74	4.07	3.25	4.81	3.75	3.83
29	9241173	99.36	36.30	48.69	18.99	4.24	2.53	3.75	2.84	2.93
30	8089212	99.28	40.10	39.58	15.49	3.94	1.91	2.80	2.17	2.23
126	11320482	99.33	41.40	41.70	15.09	3.96	3.95	5.95	4.84	4.87
132	9995574	98.63	37.20	48.68	18.83	5.30	1.74	2.56	1.99	2.03
135	10429629	99.25	41.50	40.26	14.90	3.63	3.81	5.71	4.61	4.65
136	12033742	99.30	35.60	38.59	14.99	3.94	3.67	5.51	4.62	4.61
199	12878610	99.27	41.10	39.35	14.21	3.47	5.14	7.64	6.58	6.52
200	10648988	99.29	44.30	38.85	14.13	3.45	4.53	6.73	5.62	5.62
202	10701552	99.26	31.30	36.08	14.40	3.55	2.93	4.22	3.31	3.39
423	9266639	99.36	38.60	41.76	16.25	3.95	2.66	3.84	2.85	2.97
424	5765660	99.32	37.40	47.38	18.86	4.63	0.96	1.32	0.91	0.98
425	6186518	99.27	40.70	55.59	24.40	6.29	0.48	0.74	0.59	0.59
426	9192696	99.33	37.60	60.88	25.62	6.47	0.76	1.12	0.83	0.86
427	16026797	99.32	38.30	39.40	14.56	3.46	6.07	8.72	7.26	7.30
428	11114978	99.29	33.80	42.46	15.63	4.12	2.12	3.20	2.55	2.58
429	10339748	99.33	38.90	43.14	16.38	3.95	2.53	3.67	2.88	2.94
430	9302103	99.32	39.10	46.05	17.93	4.45	2.07	3.01	2.23	2.32
431	10558625	99.31	38.70	38.17	14.47	3.69	3.41	5.16	4.18	4.21
432	9773441	99.23	44.40	40.14	14.73	3.84	3.59	5.36	4.53	4.51
433	10314096	99.29	40.40	43.33	16.54	3.95	3.21	4.80	3.79	3.85
434	9582897	99.32	38.90	39.69	15.15	3.74	2.52	3.66	2.86	2.93
435	9730444	99.55	34.40	47.83	18.05	4.49	1.42	2.10	1.65	1.69
436	10305354	99.34	40.10	41.47	15.15	3.97	2.73	4.06	3.22	3.27
437	11285384	99.43	41.40	39.43	13.77	3.22	3.11	4.68	3.68	3.74
438	9799540	99.24	31.00	49.57	19.46	4.82	1.50	2.25	1.69	1.75
439	9726089	99.54	33.90	42.99	16.14	4.11	1.68	2.46	1.93	1.97
440	10348569	99.27	40.10	44.81	16.99	4.11	3.00	4.53	3.69	3.71
441	12254838	99.24	36.90	41.61	16.33	4.47	3.68	5.59	4.57	4.58
442	15282334	99.47	32.80	47.93	19.54	4.49	3.86	5.61	4.44	4.53
443	7439038	99.32	32.20	47.58	18.71	4.72	1.16	1.78	1.39	1.41
444	10853986	99.37	37.30	38.11	14.28	3.69	3.55	5.36	4.52	4.50
445	10199293	99.38	36.70	39.04	14.66	3.66	3.26	4.88	4.03	4.04
446	6955954	99.19	39.20	45.38	17.66	4.69	1.03	1.53	1.18	1.21
447	9961642	99.32	34.20	44.23	17.43	4.48	2.41	3.71	2.92	2.96

448	10837652	99.45	34.70	44.73	17.86	4.49	2.72	4.01	3.31	3.33
449	10489725	99.27	39.20	41.52	15.65	4.13	3.56	5.44	4.53	4.52
450	9926115	99.08	40.30	45.33	16.74	4.07	3.25	4.81	3.75	3.83
451	8704272	99.49	37.40	44.80	17.64	4.21	1.49	2.17	1.67	1.71
452	8901947	99.28	40.80	42.59	16.20	4.03	2.30	3.44	2.71	2.75
453	11504452	99.21	36.90	43.02	16.38	4.31	2.43	3.86	3.31	3.26
454	6675370	99.31	37.60	51.06	21.32	6.18	1.91	2.56	1.60	1.79
455	11479510	99.50	36.60	47.12	16.82	4.12	2.75	4.09	3.40	3.40
456	13864087	99.51	34.80	38.61	14.79	4.06	3.56	5.61	4.90	4.81
457	10380460	99.39	32.90	55.37	20.52	5.00	1.89	2.90	2.41	2.40
458	11776784	99.20	32.50	50.21	17.97	4.73	2.73	4.21	3.51	3.49
459	10417023	98.88	40.30	40.38	15.28	4.77	2.60	4.15	3.50	3.46
460	10429073	99.20	32.40	52.74	18.85	4.65	2.28	3.48	2.84	2.85
461	11140661	99.38	32.70	55.62	20.06	4.71	2.46	3.64	2.86	2.92
462	12617532	99.44	38.40	40.70	15.74	3.93	3.72	5.49	4.44	4.49
463	8785900	98.83	44.10	40.66	16.38	4.82	2.53	3.85	3.16	3.17
464	11527187	99.45	33.70	55.30	22.03	5.02	2.07	3.03	2.37	2.42
465	10563184	98.91	45.50	41.32	16.26	4.74	3.36	5.07	4.11	4.14
466	9219906	98.99	41.40	48.64	19.00	5.22	1.72	2.49	1.93	1.98
467	11166203	99.41	28.10	63.81	23.35	5.44	1.53	2.23	1.66	1.73
468	11492712	99.41	26.50	65.24	25.22	5.73	1.31	1.94	1.47	1.52
469	11667551	99.51	29.60	56.47	21.16	4.97	1.75	2.60	1.99	2.04
470	10076006	99.43	31.20	62.10	22.20	5.02	1.43	2.13	1.65	1.69
471	10178051	99.51	39.30	43.09	15.99	3.68	2.42	3.49	2.65	2.74
mean	10638410. 24	99.23	38.40	46.22	17.69	4.53	3.10	4.56	3.54	3.62
mean	10194682. 18	99.34	37.07	46.47	18.00	4.52	2.58	3.86	3.17	3.19
mean	10481730. 1	99.35	36.73	44.88	17.05	4.22	2.60	3.87	3.10	3.14
StDev	1916198.7 25	0.24	4.32	6.76	2.66	0.69	1.51	2.15	1.64	1.69
StDev	2507319.5 08	0.10	4.02	8.28	4.04	1.05	1.50	2.25	1.97	1.94
StDev	1773778.7 96	0.11	3.98	6.30	2.49	0.52	1.06	1.55	1.33	1.32

Table A.6: DNA methylation levels (in %) of enzyme locations including 1000 bp up- and downstream of each locus. Cytosine contexts are combined.

Black indexes indicate hybrid lines, red maternal and blue paternal lines.

index	CMT3	MET1	AGO4	CCA1	C4H	MOS1	NPR1	ICS1
427	0.37	7.29	0.00	NaN	3.77	40.22	9.06	3.64
428	0.00	NaN	0.58	0.00	0.00	35.70	0.57	11.27
429	13.33	6.77	0.26	NaN	13.82	33.52	14.38	0.25
430	NaN	0.00	NaN	0.00	0.44	32.47	8.91	7.74
431	NaN	2.75	2.51	NaN	7.41	35.53	9.51	6.87
432	NaN	0.81	0.90	0.00	NaN	47.72	5.92	4.24
433	NaN	0.00	NaN	0.00	4.67	42.92	5.35	5.55
434	NaN	0.00	0.00	NaN	3.39	50.51	4.13	7.55
435	NaN	0.00	0.00	NaN	0.00	34.88	8.33	0.95
436	NaN	5.26	0.00	0.00	0.00	43.60	5.52	5.80
437	0.00	4.87	NaN	0.00	NaN	46.34	1.04	3.88
438	8.33	9.52	NaN	0.00	13.65	36.47	9.43	2.50
439	NaN	10.20	NaN	10.00	8.70	45.36	6.58	4.36
440	46.67	4.08	NaN	0.00	6.99	45.38	7.44	6.79
441	NaN	1.48	0.00	NaN	NaN	44.35	9.05	0.00
442	0.00	0.00	0.28	NaN	2.88	38.55	6.05	15.03
443	NaN	2.75	NaN	NaN	NaN	50.17	13.95	19.53
444	0.00	0.42	2.22	NaN	0.00	40.17	9.28	5.89
445	11.67	6.22	NaN	1.03	5.56	41.00	9.10	17.39
446	0.00	NaN	0.00	NaN	0.00	52.17	6.67	3.32
447	NaN	0.25	0.00	0.00	10.00	39.98	7.75	13.33
448	21.01	0.00	NaN	10.00	NaN	41.52	14.89	0.00
449	20.00	1.64	0.00	NaN	NaN	40.76	11.38	3.34
450	0.00	8.38	NaN	0.00	NaN	47.66	5.75	1.18
451	0.00	0.00	0.00	NaN	1.50	48.28	20.28	1.67
452	NaN	NaN	0.69	NaN	0.00	40.75	6.33	4.19
453	NaN	NaN	0.00	NaN	NaN	45.89	NaN	NaN
468	NaN	NaN	NaN	NaN	NaN	49.31	0.00	NaN
469	NaN	NaN	NaN	NaN	NaN	40.85	NaN	NaN
470	NaN	NaN	NaN	NaN	NaN	49.39	NaN	0.56
471	4.00	8.61	0.00	NaN	27.78	34.67	6.00	8.44
199	49.21	14.60	NaN	0.57	0.00	48.61	3.53	8.44
200	NaN	0.70	NaN	0.87	NaN	33.29	2.25	3.60
202	NaN	NaN	4.33	0.00	5.07	28.31	9.13	5.16
423	NaN	5.81	0.00	0.00	0.26	44.00	3.70	1.58
424	NaN	0.95	0.00	NaN	0.00	0.00	3.79	6.67
425	NaN	NaN	NaN	NaN	NaN	45.33	NaN	NaN
426	NaN	5.00	0.00	NaN	NaN	39.07	NaN	0.00
455	47.93	0.00	0.00	NaN	13.15	39.14	3.18	0.43
456	NaN	NaN	14.25	NaN	NaN	45.30	NaN	NaN
457	NaN	NaN	NaN	NaN	NaN	43.96	0.42	NaN
458	NaN	NaN	NaN	NaN	NaN	43.40	4.60	5.82
9	NaN	21.56	0.00	0.00	0.00	22.50	9.90	0.09
11	NaN	0.46	0.00	0.00	0.61	40.00	11.36	0.00
17	8.72	5.88	0.00	0.00	1.79	8.35	7.76	2.51

18	NaN	5.97	0.12	0.00	3.90	46.13	7.85	6.10
25	0.00	8.38	NaN	0.00	NaN	47.66	5.75	1.18
29	NaN	0.00	0.00	NaN	0.00	51.92	3.01	0.92
30	4.63	1.83	0.95	NaN	2.00	38.45	8.81	9.51
126	NaN	5.29	0.00	NaN	0.00	45.56	5.69	2.63
132	0.00	0.83	NaN	0.00	1.59	40.95	16.78	21.09
135	52.38	8.65	NaN	NaN	0.00	45.62	5.12	0.77
136	20.00	4.29	0.78	NaN	6.80	46.75	14.03	2.10
454	NaN	0.00	0.00	4.17	NaN	3.33	0.90	3.10
459	NaN	NaN	49.22	NaN	NaN	44.39	NaN	NaN
460	NaN	3.78	NaN	NaN	NaN	46.58	6.35	NaN
461	NaN	NaN	NaN	0.00	NaN	38.63	15.03	NaN
462	NaN	6.79	1.03	NaN	5.63	41.84	12.00	6.63
463	NaN	NaN	NaN	NaN	NaN	42.46	0.00	NaN
464	0.00	NaN	NaN	NaN	0.00	38.69	4.28	0.51
465	0.00	NaN	NaN	NaN	NaN	44.10	0.85	27.50
466	0.00	1.47	NaN	NaN	3.33	43.20	3.19	0.39
467	NaN	NaN	NaN	NaN	0.00	22.22	16.89	1.48

Table A.7: mRNA expression levels (in FPKM) of enzyme locations including 1000 bp up- and downstream of each locus. Black indexes indicate hybrid lines, red maternal and blue paternal lines.

index	CMT3	DRM2	DDM1	AGO4	CCA1	C4H	NPR1	ICS1	ROS1
427	2.04	0.61	32.24	5.49	2.48	4.48	0.88	7.03	16.20
428	1.58	2.14	30.02	7.05	2.47	5.01	NaN	4.69	21.41
429	2.05	NaN	36.74	4.62	2.48	3.56	0.76	4.91	14.26
430	1.44	1.08	33.97	5.50	1.97	4.72	1.79	3.45	21.73
431	1.22	0.87	42.46	3.82	3.72	4.47	1.51	5.67	15.90
432	1.51	NaN	33.81	6.82	3.01	4.20	1.99	4.37	16.06
433	0.88	1.38	33.81	4.83	2.21	4.43	1.24	4.76	19.38
434	1.13	0.94	38.81	3.22	2.80	4.67	0.76	5.15	18.62
435	NaN	NaN	37.61	7.43	2.53	3.72	0.79	3.96	17.47
436	NaN	1.64	34.57	NaN	3.23	3.85	1.44	5.81	16.86
437	1.37	1.36	34.09	2.38	NaN	6.04	NaN	4.95	18.52
438	1.99	0.62	35.59	3.16	2.15	3.43	1.46	4.23	16.28
439	0.91	NaN	38.59	6.49	3.29	5.13	NaN	5.46	26.65
440	2.41	NaN	34.34	2.98	NaN	4.16	1.11	7.11	15.04
441	1.53	NaN	31.93	6.43	3.24	3.17	1.25	4.98	31.23
442	1.93	NaN	33.04	2.88	NaN	4.60	1.61	4.48	30.78
443	1.13	1.11	36.15	2.43	3.17	4.12	1.26	5.70	35.67
444	1.99	NaN	31.87	6.57	NaN	4.27	NaN	6.76	34.64
445	2.23	0.68	38.34	2.35	1.61	4.23	0.81	4.65	23.81
446	NaN	1.22	35.56	NaN	NaN	4.27	NaN	5.09	42.43
447	2.60	0.94	32.74	7.21	2.78	4.20	1.21	4.21	31.41
448	1.83	NaN	34.77	6.39	NaN	4.88	1.75	4.76	24.58
449	1.11	1.25	39.37	4.17	0.87	4.35	NaN	5.60	37.61

450	1.60	NaN	28.72	2.05	2.61	4.44	1.95	5.38	21.61
451	2.51	0.70	33.03	4.10	NaN	5.59	0.71	5.84	NaN
452	NaN	0.94	36.35	6.17	2.49	4.21	1.22	6.16	25.84
453	2.57	0.81	31.73	5.78	3.00	3.83	1.22	4.83	20.62
468	1.54	1.09	31.95	2.96	1.32	4.63	NaN	5.74	19.63
469	1.68	NaN	27.12	4.57	2.47	4.75	2.38	3.84	13.09
470	0.91	NaN	34.67	4.94	1.67	5.33	0.82	4.78	15.57
471	1.51	NaN	35.06	3.66	0.94	5.34	1.19	3.68	16.07
199	2.26	1.57	41.23	9.80	NaN	5.72	2.06	5.62	15.23
200	NaN	2.99	43.62	10.97	3.89	4.34	2.19	5.46	15.42
202	1.89	4.36	20.74	6.95	4.66	4.35	1.18	5.18	34.52
455	NaN	1.04	33.58	5.91	2.08	4.67	1.73	5.43	13.25
456	2.16	0.77	31.12	2.13	1.77	5.98	0.81	6.13	20.06
457	1.91	NaN	29.72	4.24	1.92	4.49	2.61	6.14	28.80
458	2.26	0.64	34.97	3.56	NaN	4.63	NaN	5.37	20.73
9	NaN	4.14	36.77	11.67	NaN	5.31	1.36	7.25	20.64
11	1.68	3.07	28.03	6.67	6.01	5.12	1.16	8.85	17.73
17	2.21	0.76	33.21	3.60	NaN	5.32	1.88	4.08	23.66
18	NaN	1.13	36.83	9.49	NaN	4.29	1.58	6.20	13.46
25	2.03	3.20	34.92	9.17	4.16	4.78	1.38	8.43	19.85
29	2.27	NaN	25.57	8.85	7.85	4.99	NaN	5.07	16.25
30	1.91	1.28	35.40	6.90	6.15	4.54	1.04	7.19	20.21
126	2.45	1.90	26.54	9.33	3.33	3.89	NaN	4.71	12.21
132	2.19	0.90	32.41	7.59	2.59	5.02	1.66	6.65	NaN
135	1.66	2.80	35.00	9.98	NaN	5.06	NaN	6.08	17.37
136	2.74	2.35	27.25	3.91	5.77	3.84	1.69	7.05	11.47
454	1.15	1.41	35.37	5.57	4.08	5.11	NaN	5.09	NaN
459	2.46	1.11	41.27	5.04	4.09	4.18	NaN	6.40	NaN
460	1.67	NaN	39.46	4.71	1.76	4.41	NaN	6.07	31.75
461	3.34	1.52	34.25	4.75	1.73	2.93	0.74	5.98	32.57
462	NaN	1.61	33.93	2.68	NaN	4.70	1.50	5.54	28.57
463	2.38	1.87	35.45	4.68	1.90	3.29	NaN	5.59	16.46
464	1.91	0.94	37.03	6.03	3.06	4.41	NaN	6.15	16.63
465	1.61	NaN	45.19	2.04	NaN	4.41	1.81	5.62	39.73
466	2.22	0.69	30.61	5.36	0.97	4.53	1.07	6.51	NaN
467	3.25	NaN	28.32	4.03	NaN	4.17	1.61	7.40	0.87

Table A.8: Location of heterosis-associated DMRs per context.

chromosome	Start	Stop	association	Present in hybrid lines
CpG DMRs				
chrC04_random	3375823	3375881	negative	430, 431, 432, 435, 436, 437, 442
chrC03_random	4823969	4824027	positive	427, 430, 434, 446, 447, 448, 450
chrC01	24494997	24495058	negative	431, 432, 435, 436, 437, 451, 452
chrC04	10801835	10801856	negative	430, 434, 435, 437, 441, 442, 451, 452, 453
chrC04	44181803	44181865	negative	428, 432, 433, 434, 435, 436, 437
chrAnn_random	32935528	32935536	negative	428, 431, 432, 434, 436, 437

chrC07	16188782	16188793	positive	427, 428, 429, 434, 446, 450
chrC09	25187165	25187206	negative	430, 431, 434, 435, 436, 437
chrC01	18357681	18357742	negative	431, 432, 434, 435, 436, 437
chrC04	43640978	43641019	negative	432, 434, 435, 436, 437, 453
chrA08	16261351	16261417	negative	431, 432, 434, 435, 436, 437
chrA05	18446835	18446869	negative	428, 430, 436, 437, 438, 441
chrAnn_random	896770	896817	negative	431, 434, 435, 436, 437, 441
chrA02	2393040	2393083	negative	428, 431, 432, 435, 437
chrC07_random	700019	700065	negative	428, 431, 432, 434, 435, 437, 438, 453
CHG DMRs				
chrAnn_random	13196369	13196411	negative	428, 432, 434, 435, 437, 442, 451
chrAnn_random	41898686	41898694	negative	431, 432, 434, 435, 436, 437
chrAnn_random	13956106	13956146	negative	431, 432, 436, 437, 451
chrC07	6783489	6783502	negative	431, 432, 435, 436, 437,
chrC01	30226841	30226875	negative	431, 432, 435, 437, 451
chrC04	15769582	15769635	negative	431, 432, 433, 434, 435, 436, 437, 441
chrC03	44139872	44139965	negative	427, 428, 430, 431, 432, 433, 441, 442
CHH DMRs				
chrA05	16374265	16374273	negative	431, 431, 435, 436, 437

Table A.9: Mean methylation rates given in literature of *Brassica napus* and its ancestors *Brassica rapa* and *Brassica oleracea*.

	tissue	<i>Brassica</i>								
		<i>oleracea</i>			<i>rapa</i>			<i>napus</i>		
		CpG [%]	CHG [%]	CHH [%]	CpG [%]	CHG [%]	CHH [%]	CpG [%]	CHG [%]	CHH [%]
Niederhuth et al. (2016)	leaf	55.0	25.0	6.0	40.0	20.0	5.0			
Parkin et al. (2014)	leaf	54.9	9.4	2.4						
Chen et al. (2015)	young leaves from tissue culture				52.4	31.8	8.3			
Shen et al. (2017)	young flower buds							59.3	26.3	11.3
this study	whole seedling							36.1 - 65.2	13.8 - 25.6	3.2 - 6.5

Declaration on oath / Eidesstattliche Erklärung

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than the acknowledged resources and aids.

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Stuttgart, date / Stuttgart, den

Signature / Unterschrift

Confirmation of correct English




**ST PETER'S
COLLEGE**
UNIVERSITY OF OXFORD

To whom it may concern,

I am writing to confirm that I have read the thesis submitted by Susanne Edelmann. I can vouch for the high quality of English throughout the text and I can confirm that the work in this thesis can be understood easily.

Yours faithfully,


(Dr. Robert Grant-Downton)

Stipendiary Lecturer in Biological Sciences,
St. Peter's College,
University of Oxford,
New Inn Hall Street,
Oxford OX1 2DL
England

Acknowledgement

I would especially like to thank Prof. Dr. Stefan Scholten for providing me the chance to conduct my PhD studies on these fascinating projects, which offered me the opportunity to gain experience in multiple areas including developmental biology, bioinformatics and plant breeding. I furthermore appreciate his patience and support at any time if problems or questions arose and providing high performance IT equipment.

I would like to thank Prof. Dr. Arp Schnittger to agree to examine my thesis and the friendly and open affiliation of our lab group.

I want to express my special thanks to Dr. Felix Seifert, who not only introduced me very patiently to bioinformatics, but also always provided great help and support at any time and concerning any question. I would further like to thank you him for the detailed proof reading of my thesis and providing the processed sRNA data. I really enjoyed the BBQs and get-togethers that we had with the Hamburg lab group.

I would like to thank Christian Rockmann for his help and support in bioinformatical analysis of the data, for providing processed mRNA data and for being a great collaboration partner, with whom I enjoyed Greek and Mexican food after work.

Additionally, I thank all collaboration partners from the PROGRess project, especially the Jülich Plant Phenotyping group for providing the phenotypic seedling data.

I would like to thank all former and current lab members from the University of Hamburg for their support and a nice time together. Moreover, I want to express many thanks to my current colleagues from the Plant Breeding Institute from the University of Hohenheim for their warm welcome and happy after work get-togethers in Stuttgart. Thank you also for introducing me to the world of plant breeding.

I would like to thank Dr. Robert Grant-Downton for proof-reading of my thesis.

I am deeply grateful to my family and friends supporting and encouraging me throughout the thesis. Finally, I would like to thank you, Tobi, for your unconditional and endless support during both enthusiastic and stressful times as well as the multiple times we pondered together over Python programs after work.