UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Phenotypic variation in plants

Roles for epigenetics

Lauss, K.

Publication date
2017
Document Version
Other version
License
Other

Link to publication

Citation for published version (APA):
Lauss, K. (2017). *Phenotypic variation in plants: Roles for epigenetics*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:27 Apr 2025

Chapter 4

Molecular insights into four Arabidopsis epiHybrids and their parents

Kathrin Lauss[1], Rurika Oka[1], René Wardenaar[2,3], Damar T. Anggorro[1], Frank Becker[4], Joost J.B. Keurentjes[4], Frank Johannes[2,3], Maike Stam[1]

1 University of Amsterdam, Swammerdam Institute for Life Sciences, Science Park 904 1098XH Amsterdam, The Netherlands.
2 Population epigenetics and epigenomics, Department of Plant Sciences, Technical University Munich, Liesel-Beckmann-Str. 2, 85354 Freising, Germany
3 Institute for Advanced Study, Technical University Munich, Lichtenbergstr. 2a, 85748 Garching, Germany
4 University of Wageningen, Laboratory of Genetics, Droevendaalsesteeg 1, 6708PB Wageningen, The Netherlands.

Abstract

In Chapter 3 of this thesis, we used epigenetic hybrids (epiHybrids) to demonstrate that epigenetic variation in parental lines is sufficient to cause F1 heterosis (i.e. the superior performance of a hybrid compared to its parents) in several traits in Arabidopsis. In addition we associated our heterotic phenotypes in flowering time, leaf area and plant height with three QTL^{epi} regions. Here, we aim to understand the molecular basis of these heterotic effects by analyzing the transcriptomes and methylomes of four selected epiHybrids and their parental lines.

Our analysis revealed widespread non-additive RNA expression changes, particularly at transposons, and suggests genome-wide methylome interactions in the form of trans-chromosomal (de)methylation (TC(d)M) events. We also performed preliminary analyses of the identified QTL^{epi} regions, revealing non-additive gene expression and DNA methylation changes. Which and how genes within the QTL^{epi} regions contribute to trait heterosis remains elusive though.

Introduction

Heterosis, or hybrid vigor, refers to the phenomenon of a hybrid outperforming its parents in one or more traits [63,67]. Plant traits affected can be, for example, yield, stress tolerance or biomass. Various hypotheses, often based on allelic or non-allelic gene interactions [69,70,73–75], have been proposed to explain heterosis. Rather recent findings indicated that, besides differences in genetic background, also differences in the epigenomic background contribute to heterosis [72,103,132]. Changes in levels of small regulatory RNAs and/or DNA methylation were observed in Arabidopsis, maize, rice and tomato hybrids, relative to the parental lines [45,47,76,77]. The parental lines used for these studies differed in both their genetic and epigenetic profile, making it challenging to disentangle genetic from epigenetic effects. To

address the contribution of epigenetic mechanisms to heterosis independently of genetic variation, Arabidopsis lines from two different populations of epigenetic recombinant inbred lines (epiRILS) [52,53] were used to generate epigenetic F1 hybrids (epiHybrids) [58]; Chapter 3). Both epiRIL populations are near isogenic, but their genomes are a mosaic in terms of epigenetic patterns. The two populations have been created by crossing wild-type Columbia (Col-wt) with Col-wt lines carrying a mutation in either *METHYLTRANSFERASE 1 (MET1-3*) or *DECREASE IN DNA METHYLATION 1* (*DDM1–2*).

DNA-METHYLTRANSFERASE1 (MET1) is involved in maintenance of DNA methylation at cytosines in CG sequence context, but *met1* mutants also show abnormalities in non-CG methylation and the repressive histone modification dimethylation of histone H3 Lysine 9 (H3K9me2) [25,54]. More specifically, loss of MET1 causes an almost complete elimination of CG methylation at genes and TEs, a substantial ectopic CHG and CHH hypermethylation at thousands of genes and TEs [54,133,134], and is linked to misregulated expression of affected genes and transcriptional activation of TEs [133,135–137]. However, despite transcriptional activation, many autonomous TEs remain immobile [138,139].

Loss of DDM1 results in a ~ 70% reduction in DNA methylation [26], predominantly affecting CG and CHG methylation (57 % and 58% reduction) and to a lesser extent CHH methylation (32% reduction) [18]. The methylation loss concerns mainly transposable elements, among which the major targets appear to be long TEs [18]. Loss of DDM1 also affects genic loci where it reduces CG methylation and causes CHG hypermethylation in gene bodies [18].

In an epiHybrid generated by crossing a *met1*-derived epiRIL with Col-wt, heterosis for biomass was reported [58]. The effect was only observed with the epiRIL as maternal parent and not in the reciprocal cross, suggesting a strong influence of the maternal cytoplasm [58]. In chapter 3

of this thesis, we created epiHybrids by crossing Col-wt as a maternal parent to nineteen different *ddm1*-derived epiRILs. We observed various positive and negative heterotic effects in one or more of the six traits monitored among the nineteen epiHybrids, indicating that epigenetic divergence among parents can be sufficient to cause heterosis.

Both the *met1* and *ddm1* epiRIL populations consist of several lines that each have a different DNA methylation pattern. Divergence in DNA methylation patterns can affect the activity of (non-)genic regions depending on the location of the modification. DNA methylation at genes can be present in both promoter regions and the gene body [140]. Promoter methylation commonly results in reduced gene transcription while gene-body methylation, which predominantly occurs on long, evolutionary conserved genes, is associated with moderate-to-high gene expression [22,23]. There is a lot of variation in DNA methylation patterns, at both genes and transposable elements, across natural Arabidopsis populations [140] and among subspecies of maize and rice [141–144]. Insights into how this variation affects plant phenotypes and adaptation to local environments remain scarce.

The contribution of DNA methylation (or other epigenetic modifications) to natural variation in plant traits is exemplified by a range of plant phenotypes that are caused by epialleles [34,35,105]. Epialleles are allele variants with the same DNA sequence but different epigenetic profiles and gene expression levels. To date, epialleles have been described in multiple plant species, e.g. Arabidopsis, maize, rice and tomato, where they affect traits like flowering time, plant pigmentation or vitamin E accumulation [36–40]. Differential DNA methylation at epialleles can affect promoter regions and/or gene bodies [38,40] but can also concern regulatory elements (e.g. enhancers) [37,39]. Most described epialleles are associated with DNA methylation differences [34,35], which is often

interconnected with the repressive histone modification H3K9 methylation [15,16,145].

Variation in the level of histone modifications not associated with DNA methylation can also be related with differences in phenotypes. This is exemplified by a study showing that altered profiles of activating histone modifications at genes affecting the circadian rhythm in an Arabidopsis allotetraploid (*A.thaliana* x *A.arenosa*) are associated with increases in chlorophyll, starch and sugar content in these plants [120].

During hybridization between two different parental lines, *in trans* communication can occur between homologous, differentially or similarly methylated regions (DMRs). When *in trans* communication between low and high methylated regions results in the low methylated one acquiring DNA methylation this behavior is reminiscent of paramutation [46,79], a phenomenon resulting in heritable changes in epigenetic profiles [80]. Paramutation has been extensively studied at particular loci in maize, where it causes non-Mendelian inheritance patterns of traits [81]. Similar events were reported on a genome-wide scale at various DMRs in Arabidopsis hybrids between the accessions C24 and Ler [79]. These events were termed trans-chromosomal methylation (TCM). The same study also showed that interactions between DMRs cannot only result in a gain but also in a loss of DNA methylation at affected sequences. Such loss is called trans-chromosomal demethylation (TCdM) [79]. A selection of the TCM methylation patterns were examined in more detail and were associated with decreases in gene expression and the activating histone modification H3K9 acetylation [121]. Also, some of the examined TC(d)M cases were stably inherited to the next generation [121]. The mechanism underlying these methylome interactions appears to rely on RNA-dependent DNA methylation (RdDM), a pathway involving small RNAs that induce DNA methylation at homologous target sequences [17,146]. In line with that, diffusible small RNAs have been proposed to be involved in

TC(d)M [46,79,146]. Genetic variation at the regions undergoing such events was positively associated with TCdM and negatively associated with TCM, implying that small RNAs derived from genetically divergent alleles are not proficient in inducing *in trans* methylation while small RNAs derived from homologous regions are [146].

A recent paper investigated genome-wide changes in DNA methylation and transcript levels in an epiHybrid between Col-wt and an isogenic *met1* mutant line [78]. In this epiHybrid, heritable epialleles were arising at many genic loci but most changes in DNA methylation were shown to be restricted to pericentromeric transposons (TEs). Those changes included both increases and decreases of DNA methylation at subsets of TEs. Interestingly, the study described one example of a trans-demethylation event in the F1 being associated with TE-mobilization (of an EN-SPM element) in the F2, demonstrating that hypomethylation occurring in the F1 can, in the F2 progeny, trigger mobilization of a TE that has was immobile in the parents. The authors also described the spontaneous emergence of a heritable epiallele, displaying a non-parental methylation pattern, at the *INCREASE IN BONSAI METHYLATION 1* (*IBM1*) gene in the F1 epiHybrid, showing that novel, meiotically heritable, methylation patterns can be created during hybridization.

For understanding whether and where DNA methylome patterns contribute to quantitative traits and phenomena like heterosis, methylation profiles at particular loci need to be linked to mRNA levels and to phenotypes, for example, by mapping epigenetic quantitative trait loci ($QTLs^{epi}$). Recent approaches using *ddm1*-derived epiRILs already identified $QTLs^{epi}$ in Arabidopsis for flowering time, root length and plasticity [56,57]. In chapter 3 of this thesis, we identified $QTLs^{epi}$ for heterosis in flowering time (FT), leaf area (LA) and plant height (HT) that were associated with differentially methylated regions (DMRs) in the parental genomes of our epiHybrids. These DMRs contained 35 genes for

the FT-LA DMRs and 14 genes for the HT-DMRs that potentially underlie the heterotic phenotypes observed.

Here, we wanted to get a better understanding of the molecular mechanisms behind the heterotic phenotypes effects. To that end we analyzed the transcriptomes and methylomes of four selected epiHybrids and their parental lines. We found widespread non-additive transcriptional changes, particularly at transposons, and strong indications for global methylome interactions in the form of trans-chromosomal (de) methylation (TC(d)M) events. We also performed more detailed preliminary analyses on DNA methylation and RNA expression at genes within the identified QTL^{epi} intervals revealing non-additive gene expression and DNA methylation changes.

Results

Experimental Approach and features of parental lines

The epiRILs used for our study originate from a cross between Col-wt and a *ddm1*-mutant parent (in Col-wt background), followed by backcrossing to Col-wt, selection of plants wildtype for DDM1, and several rounds of selfing. This subsequently resulted in the *ddm1*-derived epiRIL population [52]. The epiRIL genomes are a mosaic of Col-wt-derived regions (with wt DNA methylation profile) and *ddm1*-derived regions (with *ddm1* DNA methylation profile) in an epi-homozygous state [52]. Loss of DDM1 causes transcriptional activation of TEs and genes in and around pericentromeres [18,52].

In Chapter 3 we crossed Col-wt to a selection of epiRILs to create epiHybrids, which we subsequently screened for heterotic phenotypes. Here, we performed DNA methylome and RNA expression analysis on four of those epiHybrids (92H, 150H, 193H and 232H) and their parental lines. The four epiHybrids were selected based on their heterotic phenotypes in traits we previously identified putative $QTLs^{epi}$ for: Flowering time (FT),

rosette leaf area (LA) and plant height (HT) (Chapter 3). All four epiHybrids displayed negative heterosis in FT: 92H, 150H, 193H showed negative mid-parent heterosis (MPH) and 232H Low-parent heterosis (LPH). For LA, 92H displayed an additive effect, 150H positive MPH whilst 193H and 232H performed better than their best parent (BPH). And for HT, 92H and 232H showed additive expression, 150H positive MPH and 193H BPH.

We performed whole genome bisulfite-sequencing on 21DAS old aerial rosette tissue from the four epiHybrids and their parental lines. For each line, we examined two biological replicates, which displayed a good correlation (r_s > 0.8; SI Fig.1); the methylomes were subsequently combined for further analysis. Similarly, we performed RNA-sequencing on 21DAS old aerial rosette tissue of the four epiHybrids and their parental lines. We here examined two or three biological replicates (*see methods*) for each line. For analysis, only uniquely-mapped reads were retained. The correlation between the biological replicates was good, therefore reads from replicate samples were pooled for further analysis (r_s > 0.9; SI Table 1). The methylation analysis was done with 1 kb sliding windows; steps of 500bps (*see methods*).

As intended, the regions in the epiRILs originating from *ddm1* mostly exhibited strongly reduced methylation as compared to the same regions in the Col-wt, where the reduction in methylation was the strongest in the pericentromeric regions (Fig. 1). When comparing RNA expression of the epiRIL parental lines with that of Col-wt, divergence of expression was observed along the entire chromosomes At the noncentromeric regions the divergence was both positive and negative. The positive mRNA divergence was particularly strong at some pericentromeric regions, where this did correlate with decreased methylation (Fig. 1). For instance, the pericentromeres of chromosome 3 in line 92 and 150 consist mostly of Col-wt-derived regions, and there were only marginal differences in RNA expression and DNA methylation between these regions in the epiRILs and

Col-wt (Fig. 1). In lines 193 and 232, the pericentromeres of chromosome 3 are *ddm1*-derived, which is reflected in an decrease in DNA methylation divergence and increase in RNA expression between those epiRILs and Col-wt (Fig. 1). Lines 92 and 232 are the epiRILs with most pericentromeres (four and five, respectively) derived from *ddm1* (Fig. 1).

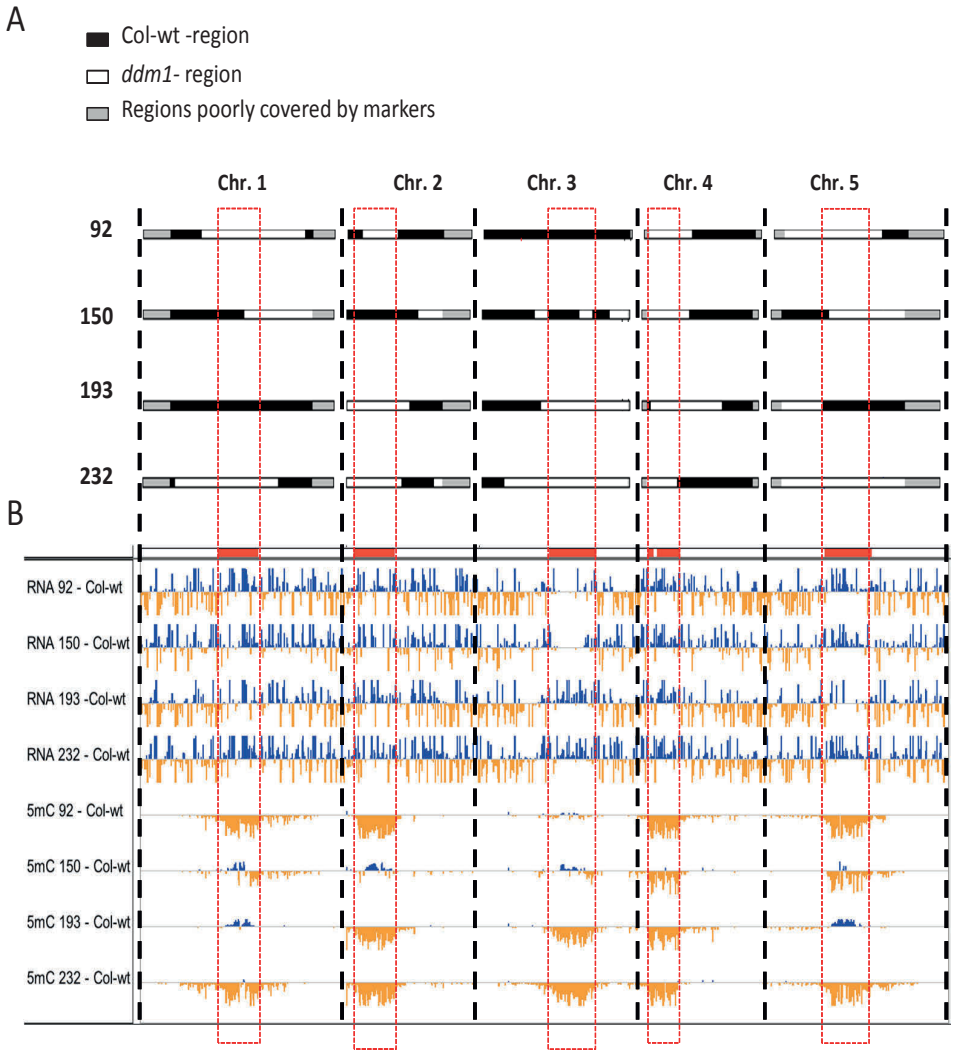


Figure 1: RNA expression and DNA methylation (5mC) analysis comparing epiRIL parents vs Col-wt. A) Schematic depiction of the recombination map in epiRIL parents with Col-wt-derived regions (black), *ddm1*-derived regions (white) and regions poorly covered by markers (grey) indicated. B) Displays log2-fold expression differences (range -10 to 10) and DNA methylation differences (range -0.1 to 0.1) between indicated epiRIL parents and Col-wt (wt) along the five Arabidopsis chromosomes. Blue indicates positive differences and orange negative differences. Pericentromeres are indicated in red. The dashed red boxes indicate the pericentromeric regions in A) and B).

High variation in DNA methylation in epiHybrids at regions originating from the *ddm1*-mutant

Re-shaping of the methylome during hybridization has been proposed to be involved in heterosis in genetic hybrids, and was implicated in the formation of novel epialleles in a *met1*-derived epiHybrid [47,48,78,79]. Here we examined the methylome changes that occurred in four epiHybrids generated using *ddm1*-derived epiRILs (Chapter 3). The genomes of epiHybrids that were created by crossing Col-wt with epiRIL lines, consist of epi-homozygous Col-wt-regions (from hereon called Col-wt regions) and epi-heterozygous *ddm1*-derived regions (from hereon called *ddm1* regions). In order to assess whether there are non-additive methylation alterations (" methylation heterosis") in the epiHybrids, we started out analyzing non-additive methylation changes at *ddm1*- and Col-wt-derived regions of the epiHybrids.

As expected, overall DNA methylation levels were higher in Col-wt than in any epiRIL parent (SI Fig. 2). The overall methylation levels of the epiHybrids were in between their parental lines, indicating an additive effect on global methylation levels (SI Fig. 2). We next investigated non-additive methylation changes in the epiHybrids, for Col-wt windows and *ddm1* windows separately. In all four epiHybrids when comparing *ddm1*-derived windows with Col-wt-derived windows we observe more variance in windows originating from *ddm1* (Fig. 2). We observed the highest

variation in methylation divergence from mid-parent value (MPV) in CG sequence context, followed by CHG and finally CHH sequence context (where H stands for non-G nucleotides, Fig. 2). But non-additive methylation changes occurred in all three contexts.

In epiHybrids, hypomethylated *ddm1*-derived regions are combined with Col-wt regions, resulting in epi-heterozygosity. Homologous sequence regions displaying different, epigenetic profiles have been shown to undergo non-additive methylation changes, also called trans-chromosomal (de)methylation [46,78,79]. In a *ddm1*-mutant, one of the two founding parents of the epiRILS, DNA methylation was mostly lost at CG and CHG context and least in CHH context [18]. Therefore, it was not unexpected that most non-additive changes occur at the CG and CHG sequence contexts at which methylation was lost in *ddm1*-mutants [18].

The medians of the *ddm1*-derived windows in CG context were, except for 92H, above the MPV (Fig. 2). This might reflect re-methylation events at regions with CG context, similar to what has been observed in a *met1*-derived epiHybrid [78].

Taken together, these findings strongly suggest the occurrence of trans-chromosomal (de)methylation events.

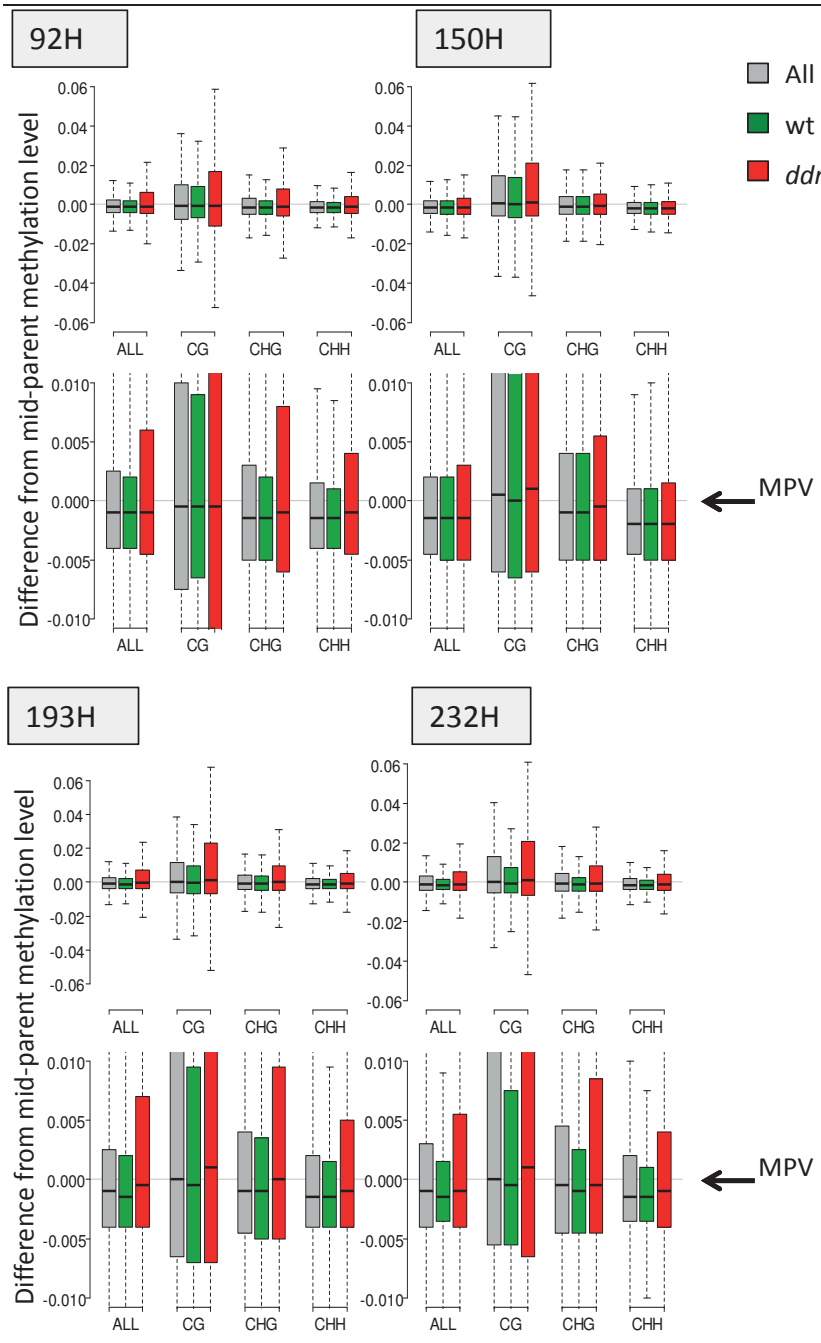

92H
150H
193H
232H
All windows
wt windows
ddm1 windows
Difference from mid-parent methylation level
0.06
0.04
0.02
0.00
-0.02
-0.04
-0.06
0.010
0.005
0.000
-0.005
-0.010
ALL
CG
CHG
CHH
MPV

Figure 2: DNA methylation divergence from mid-parent value (MPV) in the epiHybrids. Shown are all windows combined (grey), epi-homozygous Col-wt windows (green) and epi-heterozygous *ddm1* windows (red) in the different sequence contexts for the four epiHybrids. Windows refers to 1 kb sliding window, step size 500 bp throughout the genome (see Methods). DNA methylation level is determined as methylated reads per cytosine/ total reads per cytosine, averaged over the entire window. Results are shown for methylation levels that were based on all cytosines (ALL) or cytosines of one context only (CG, CHG and CHH). The bottom panel shows a zoom in from the top panel.

Trans-chromosomal (de)methylation events occur in the epiHybrids

Trans-chromosomal (de)methylation events have been described to occur in genetic hybrids as well as in a *met1*-derived epigenetic hybrid [78,79,146]. If such events result in a heritable change in epigenetic profile, they cause non-Mendelian segregation of epialleles (paramutation; [81]). These methylome interactions have been shown to be dependent on RdDM in Arabidopsis and often involve small regulatory RNAs [78,79,146].

In order to examine the non-additive methylation changes in more detail and quantify strong non-additive divergence, we analyzed the DNA methylomes of the epiHybrids and their parental lines by sliding along the genome in windows of 1 kb with a step size of 500 bp. Only windows with a minimum read coverage of 3 per cytosine for at least 50 cytosines were included. From the parental datasets we calculated mid-parent DNA methylation levels per window. Then the difference from the mid-parent value (MPV) of DNA methylation was determined for all windows of the epiHybrids. Outlier windows, meaning windows that deviated + or – 3 SD from the mid-parent levels, were identified (Fig. 3, Table 1). These "outlier" windows could represent allelic and non-allelic trans-chromosomal (de)methylation (TC(d)M) events.

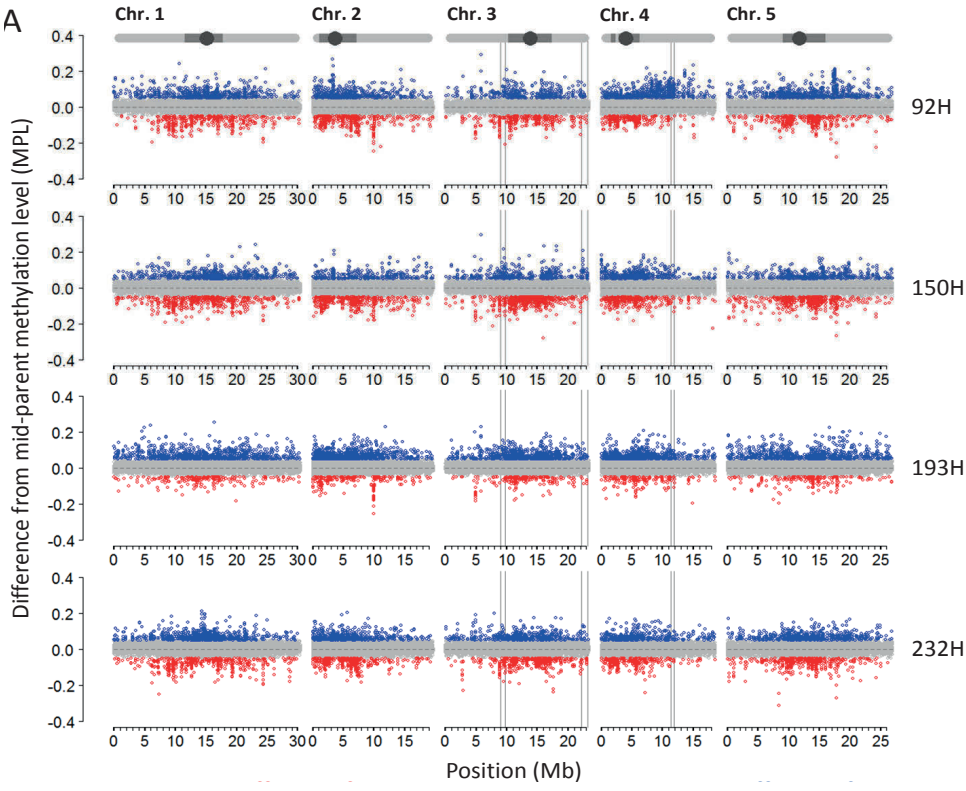
We found that 0.5 to 1.3% of all windows across the 4 epiHybrids show negative deviation from mid-parent-value, while 1.2 to 2% show positive methylation deviation from the MPV (Table 1). The difference between

the percentage ranges may indicate that a gain in methylation occurs more often than a loss. However, an amplification bias towards methylated fragments during construction of the sequencing libraries cannot be excluded [147]. Consistent with a prior report on Arabidopsis hybrids (C24 crossed with Ler), putative TC(d)M events were enriched at regions that were differentially methylated between the parents (Table 1) [79]. More specifically, 41-75% of all TCdM and 28-40% of all TCM events displayed differential methylation between the parents (Table 1). This may constitute further support for more allelic than non-allelic interactions. However, we also identified many putative TC(d)M events, particularly TCM, at regions that did not show significant methylation differences between the parents (Table 1). This is consistent with a recent report on another Arabidopsis hybrid (Col crossed with C24) that observed even more TCM events between loci that are similarly methylated in the parents than between DMRs [146].

Our detected putative TC(d)M events occurred in pericentromeric regions as well as along chromosome arms (Fig. 3A). Despite that most differences from MPV occurred at *ddm1*-derived epi-heterozygous regions, TC(d)M also appeared to occur at epi-homozygous Col-wt regions (Fig. 3A). This is an indication that TC(d)M events might also play a role at regions with very similar methylation profiles or could hint towards non-allelic TC(d)M events that have been described for transposons.

Next, we looked into whether the putative TC(d)M regions were shared among the four epiHybrids. We found only a minority of TCdM and TCM events, 85 and 104, respectively, that were shared among all four epiHybrids (Fig. 3B). About half of the TC(d)M events were unique to the individual epiHybrids (Fig. 3B). This could be interpreted as the specific DNA methylomes of the different epiRILs strongly influencing methylome interactions. It will be interesting to investigate where the shared TC(d)M

events occur and to what extent *ddm1*-derived regions in the epiRILs play a role in triggering these events.



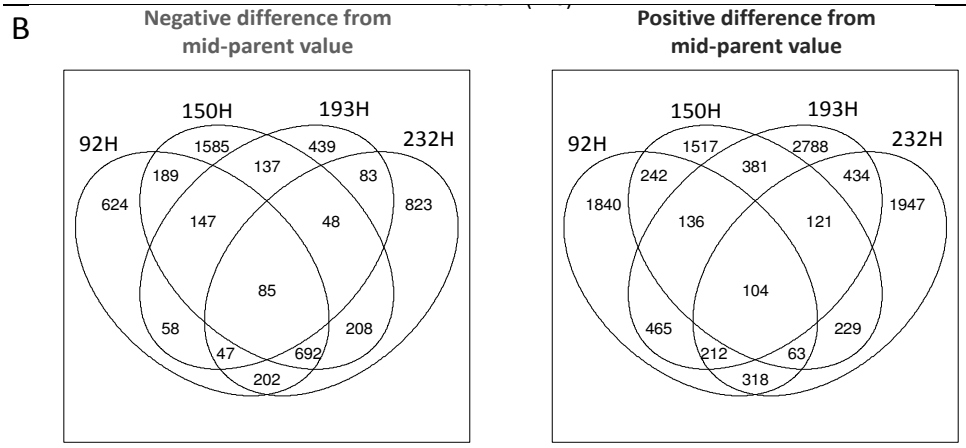


Figure 3: Detected outlier windows. (A) Difference from mid-parent value (MPV) of DNA methylation level for the five *Arabidopsis* chromosomes (chr.) in all 4 epihybrids. Outlier windows are shown in blue (positive difference from MPV) or red (negative difference from MPV). QTL regions on chromosomes 3 and 4 are indicated with vertical lines. **(B)** Venn diagrams showing the number of outlier windows (Windows with difference >= 3 SD from MPV or <= -3 SD from MPV) with negative (red) and positive (blue) deviation that are shared between the 4 epiHybrids.

Hybrid	Description	Coverage selection[1]		Outlier windows[2]		Differentially methylated between parents[3]		
		Win. #	Win. %	Win. #	Win. %	Win. #	Win. %	Out. %
	Insuff. coverage	5,143	2.16					
	Suff. coverage	**233,153**	**97.84**					
92H	Not signif.			227,729	97.67	13,565	84.42	5.96
	Signif. lower			2,044	0.88	1,519	9.45	74.32
	Signif. higher			3,380	1.45	985	6.13	29.14
	Insuff. coverage	5,066	2.13					
	Suff. coverage	**233,230**	**97.87**					
150H	Not signif.			227,346	97.48	6,201	72.16	2.73
	Signif. lower			3,091	1.33	1,279	14.88	41.38
	Signif. higher			2,793	1.20	1,113	12.95	39.85
	Insuff. coverage	5,175	2.17					
	Suff. coverage	**233,121**	**97.83**					
193H	Not signif.			227,436	97.56	9,304	83.52	4.09
	Signif. lower			1,044	0.45	516	4.63	49.43
	Signif. higher			4,641	1.99	1,320	11.85	28.44
	Insuff. coverage	4,970	2.09					
	Suff. coverage	**233,326**	**97.91**					
232H	Not signif.			227,710	97.59	16,813	85.01	7.38
	Signif. lower			2,188	0.94	1,654	8.36	75.59
	Signif. higher			3,428	1.47	1,311	6.63	38.24

[1]Sufficient coverage: Fifty or more cytosines with a minimum read coverage of three.
[2]Outlier criteria: Difference from mid-parent value should be either minus three times the standard deviation or lower (Signif. lower) or plus three times the standard deviation or more (Signif. higher).
[3]Differential methylation between parents: Difference in methylation level should be equal or lower than -0.1 methylation level (see methods). Win. %: Percentage with respect to all windows that are also differentially methylated between the parents. Out. %: Percentage of significantly differentially methylated regions with respect to all outlier windows of that category (Outlier windows).

Table 1: The number of outlier windows compared to MPV detected genome-wide. Reported are the number of windows for which 50 or more cytosines had sufficient read coverage (>= 3; Suff. coverage), the total number of outlier windows that were significantly lower (Signif. Lower) or higher (Signif. higher) than mid-parent value (MPV) and the number of outlier windows for which the

parental lines also showed differential methylation. Also are reported the number of windows for which fewer than 50 cytosines contained a sufficient coverage (Insuff. coverage) as well as the number of windows that did not show a significant difference from the mid-parent value (Not signif.). Insuff. stands for insufficient, suff. stands for sufficient.

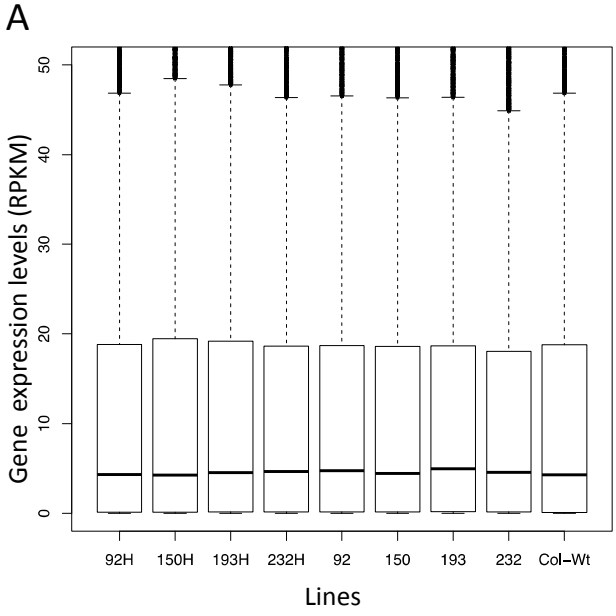
Transcription dynamics in the epiHybrids and their parents

Changes in DNA methylation can affect gene expression and TE silencing [38,40]. The epiHybrids displayed non-additive DNA methylation levels compared to the MPV, but this does not establish that they are causal for the heterosis of the traits. We wanted to investigate if the deviations in DNA methylation levels are associated with changes in transcript levels of nearby genes.

We examined overall transcript levels for coding genes and transposons in the four epiHybrids and their parental lines. Overall transcript levels of genes varied only marginally between the parental lines and the F1 epiHybrids (Fig. 4A), suggesting that the overall gene expression levels are not affected by hybridization. The expression of TEs was upregulated compared to Col-wt in all four epiRIL parents (Fig. 4B). Relatively strong upregulation was detected in epiRILs 92 and 232. The epiHybrids derived from these two lines showed a clear increase in TE expression as well, to about half the levels observed in their parental lines.

As there was no heterosis of average mRNA expression, we then asked whether individual mRNAs were subject to heterosis, i.e. differences in the hybrids from the MPVs. To determine non-additive expression changes in the epiHybrids, we used the parental datasets to calculated mid-parent value of expression per transcript and compared that to the actual expression levels in the epiHybrids. Using this approach, we indeed observed deviations from mid-parent value in terms of RNA expression and 5mC within the pericentromeric regions but also along the entire chromosomes (SI Fig. 3). One example of a region with strong expression

increases compared to MPV was found in epiHybrid 150H: as expected, these strong increases in transcript levels at the pericentromere of chromosome 3 coincided with a reduction of 5mC relative to MPV (SI Fig. S3).



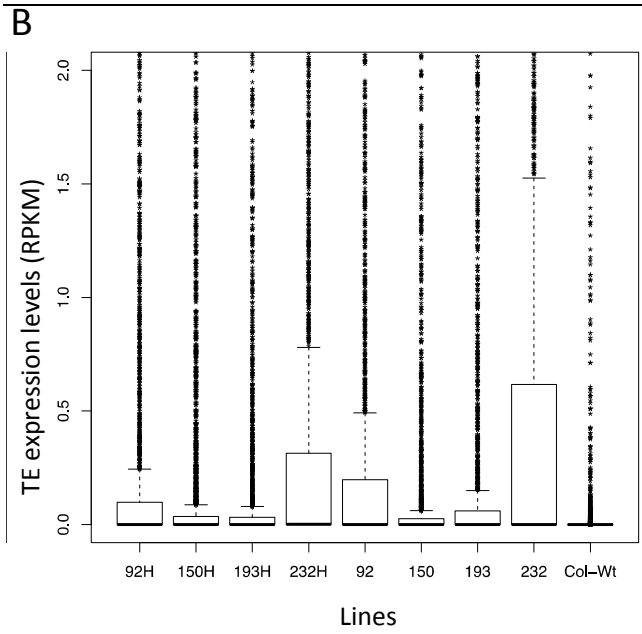


Figure 4: Global RNA expression analysis. Expression levels (uniquely-mapping reads only) of A) all coding genes and B) all transposons in epiHybrids (92H, 150H, 193H and 232H) and their parents.

Effects on transposon expression in epiHybrids

The epiRILs used in this study were derived from *ddm1*-mutant plants [52]. In *ddm1*-mutants DNA methylation can be lost at transposons, especially at heterochromatic TEs [18]. This results in transcriptional activation of TEs in *ddm1*-mutants [18]. We found increased expression levels of TEs in the epiRILs as well as in the epiHybrids, whereby epiRIL 232 and its epiHybrid showed the strongest upregulation (Fig. 4B). To investigate which TE families were responsible for the upregulation of expression of TEs, we analyzed the expression levels (uniquely-mapping reads only) of different TE families in the epiHybrids and their parental lines. We found that RathE2_cons, Line/L1, LTR/*Copia*, LTR/*Gypsy*, DNA/*MuDR*, and DNA/*En-SPM* elements are among the classes that, compared to Col-wt, were

clearly upregulated in one or more epiRIL parents and their respective epiHybrids (SI Fig. 4). In all of these TE families, 232 and 232H were among the parent-hybrid duos with the highest upregulation in various TE families (SI Fig. 4). This is in line with the highest levels of overall TE expression being displayed by epiRIL 232 and epiHybrid 232H (Fig. 4).

Next, we looked into transposons whose expression was at least 1 RPKM or higher and diverged substantially (> 2 fold) from MPV in the epiHybrids, and we discriminated between TE-reads that overlapped with annotations of protein-coding genes versus ones that do not (Table 2). We observed significant positive as well as negative divergence in expression from MPV in all epiHybrids. Positive and negative MP-divergence in the epiHybrids could again be caused by trans chromosomal methylation and demethylation events, respectively. Similar events have been described at transposons in a *met1*-derived epiHybrid [78].

LTR/*Gypsy* elements, DNA/*MuDR* and RC/*Helitrons* were among the families with the most significant cases of positive or negative divergence from MPV across the epiHybrids (Table 2). In addition, we found many cases of strong positive and negative MP-divergence in RNA expression for DNA/*EN-Spm* and LTR/*Copia* in all epiHybrids (Table 2). Despite RC/*Helitrons* having many cases of strong MP-divergence in RNA expression (Table 2), global expression levels of TEs did not appear to be strongly upregulated in any line monitored (SI Fig. 4). Possibly, this is due to only a relatively small fraction of the highly abundant RC/*Helitron* [148] family being upregulated, while the remaining RC/*Helitron* TEs were not affected.

In Arabidopsis about 36% of the genome consists of TEs, with a big fraction of them located in pericentromeric regions [30,31]. The TE families showing cases of significant divergence in expression level from MPV differed in sequence features, amplification mechanisms or preferential genomic location. LTR/*Gypsy* elements are retrotransposons with long

terminal repeats (LTRs) [18,149], while *MuDRs* are DNA transposons flanked with terminal inverted-repeats. RC/*Helitrons* are a type of DNA-transposon that amplifies via a rolling-circle mechanism [149]. The majority of DNA/*MuDR* and RC/*Helitrons* is annotated to protein-coding regions, in line with those elements being dispersed throughout the genome of Col-wt [18,148]. LTR/*Gypsy* elements, on the other hand, were mostly concentrated in the TE-rich, gene-poor pericentromeric regions [18,148] (table 2).

EpiHybrid 92H, followed by 193H, 232H and 150H, displayed the highest numbers of TEs, especially LTR/*Gypsy*, DNA/*MuDR* and RC/*Helitrons*, of which the expression strongly diverged from MPV (Table 2). When only focusing on significantly upregulated or downregulated TEs, again, 92H displayed the highest number of TEs, followed by 150H, 193H and 232H and 193H, 232H and 150H, respectively (table 2). The global TE transcript levels were highest in epiHybrid 232H (Figure 4B).

A previous study reported that a *ddm1* mutation mainly affects DNA methylation of large transposons [18]. Therefore we investigated whether the effect on TE transcript levels in epiRILs and epiHybrids is mainly observed at long TEs. When examining the size distribution of TEs per se we found that TE sizes can be very variable among TE families (SI Fig. 5), with individual TEs even exceeding the size of 4kb in DNA/*En-Spm*, DNA/*Harbinger*, DNA/*MuDR*, Line/L1, LTR/*Copia*, LTR/*Gypsy*, RC/*Helitron* and unassigned families (SI Fig. 5). The TEs of those families were largest in terms of average size (SI table 2). Next, we focused on our selection of TEs of which the expression substantially (> 2 fold) diverged from MPV in the epiHybrids (Table 2). We found the average TE size of upregulated and downregulated TEs in the different families was clearly larger than the average size of all TEs in these same families (SI table 2, SI Fig. 6). For instance, the average size of all *Gypsy* transposons was 739bp, while the

average size of the up- and downregulated *Gypsy* transposons in the epiHybrids was 2730bp and 2721bp, respectively (SI table 2).

Our results demonstrate that in general the expression of TEs is more upregulated among the epiRILs compared to Col-wt and that was reflected in the epiHybrids, in which the TE expression levels were about the average of their epiRIL parents. Expression of transposons belonging to LTR/*Gypsy*, DNA/*MuDR* and RC/*Helitron* families were most affected in the epiHybrids (in terms of significant positive and negative mid-parent divergence). In addition, our data indicated that the expression of TEs larger than average size was preferentially affected in the epiHybrids. EpiHybrids 92H and displayed most cases of significant expression divergence MPV and 232H the highest upregulation of overall TE expression. This is consistent with more of the pericentromeric regions in their parental epiRILs being *ddm1*-derived (Fig. 1).

		92H		150H		193H		232H		
		signif. up	signif. down	signif. up	signif. down	signif. up	signif. down	signif. up	signif. down	
GENES										
coding	Genes	487	623	232	154	229	231	205	233	
TRANSPOSONS										**Sum**
coding	DNA	2	4	3	0	0	1	0	0	10
coding	DNA/En-Spm	3	6	0	0	3	4	1	1	18
coding	DNA/Harbinger	1	2	0	0	0	0	0	0	3
coding	DNA/HAT	5	2	0	1	0	2	2	2	14
coding	DNA/Mariner	1	1	0	0	0	0	0	0	2
coding	DNA/MuDR	12	35	6	10	9	7	3	13	95
coding	DNA/Pogo	1	1	1	1	1	1	0	0	6
coding	DNA/Tc1	1	0	0	0	0	0	0	0	1
coding	LINE?	0	0	0	0	0	0	0	1	1
coding	LINE/L1	4	7	1	0	0	0	0	6	18
coding	LTR/Copia	7	7	0	2	2	0	0	6	24
coding	LTR/Gypsy	2	7	2	0	1	2	0	3	17
coding	RathE1_cons	1	1	0	0	1	0	2	1	6
coding	RC/Helitron	40	64	18	15	15	25	10	22	209
TE	SINE	0	2	0	0	0	0	0	2	4
TE	DNA/En-Spm	10	7	4	2	2	13	7	2	47
TE	DNA/Harbinger	1	0	1	0	0	0	1	1	4
TE	DNA/HAT	0	0	0	3	0	3	0	0	6
TE	DNA/MuDR	4	7	2	4	0	7	2	9	35
TE	LINE/L1	0	2	0	0	0	1	0	0	3
TE	LTR/Copia	4	2	1	1	0	3	1	2	14
TE	LTR/Gypsy	24	12	7	9	7	19	8	14	100
TE	RC/Helitron	0	0	0	2	0	4	0	1	7
TE	Unassigned	0	1	0	0	0	1	0	1	3
Sum		123	170	46	50	41	93	37	87	

Table 2: Genes and TEs with significant (>2 fold) expression change from mid-parent value (MPV) in positive (up) or negative (down) direction in the epiHybrids. Only uniquely mapping reads were used for the analysis. "Coding" in terms of TEs refers to TE-sequences annotated to a coding region.

Preliminary analysis on genes within the QTLs^{epi} intervals based on DNA methylation and transcription

To understand transcriptome patterns potentially relevant for our heterotic phenotypes in the epiHybrids (Chapter 3), we looked into significant gene expression changes from mid-parent value, first globally but then specifically at the previously identified QTLs^{epi}.

The total number of genes of which the expression was substantially up- or down regulated (>2 fold from MPV) varied between the epiHybrids (Table 2). EpiHybrid 92 had the most genes with differential expression from mid-parent levels with 487 genes up- and 623 genes down-regulated. For the other three hybrids those numbers were lower.

Next, we set the first steps to investigate DNA methylation profiles and gene expression in the QTLs^{epi} intervals in more detail, to see if there are links to trait heterosis (FT_1, FT_2 and HT_1, see Chapter 3). We again focused on non-additive changes of DNA methylation and RNA expression by comparing the epiHybrids with the average of their parental lines (the mid-parent value). We performed the candidate gene selection in two ways: 1) we identified genes based on MP-divergence in DNA methylation profile, followed by examining expression levels of some candidate genes and conversely 2) we identified genes in the QTLs^{epi} intervals showing significant MP-divergence in RNA expression, followed by examining DNA methylation divergence at some candidates.

We first focus on approach 1. In order to refine the selection of candidate genes and select potential candidate windows in the QTLs^{epi} intervals, we applied several filters to the outlier windows established by the DNA methylation data (windows refers to 1kb sliding window with a step size of

500bp throughout the genome; see methods). These outlier windows were considered candidate windows when they were located within one of the three QTL intervals and overlapped with the promoter or body of genes. Promoters were here defined as the 1.5kb region upstream of the transcription start site (TSS), and gene bodies were defined by TAIR (TAIR 10 release). Additional criteria were for instance that outlier windows should show significantly different methylation levels between the two parents. Also the different cytosine sequence contexts were considered in the analysis (see methods; SI table 3).

Our selection resulted in 14 genes (ten in FT QTL^{epi} 1, three in FT QTL^{epi} 2 and one in the HT QTL^{epi}) (SI Table 3). We narrowed this selection down to five genes by choosing windows with mid-parent methylation divergence in the four epiHybrids matching the differences between their parents. Here we shall further focus on the five genes covered by these windows located in the two FT $QTLs^{epi}$ regions (AT3G25480, AT3G25490, AT3G26480, AT3G61470 and AT3G61480). However, even for those likely candidates, we did not find a conclusive link with RNA expression that would prove their involvement in the regulation of trait heterosis.

AT3G25480, coding for a Rhodanese cell cycle phosphatase, displayed significant positive MP-divergence for methylation at the promoter in 150H and 232H. Gene expression levels were, however, around MP-level for 150H, 193H and 232H, while 92H showed above mid-parent expression (Fig. 5A). The higher expression in 92H was associated with a lower increase in DNA methylation compared to MPV than observed for the other three hybrids. AT3G25490, encoding for a protein kinase, displayed significant increased methylation levels at the promoter region for 150H and 232H, and positive, but not significant, MP-divergence in 92H and 193H. The gene was, however, not expressed in any of the four hybrids or their parental lines (Fig. 5B, SI Fig 7).

AT3G26480, which encodes a Transducin/WD-40 family protein, had significant negative MP-divergence in DNA methylation at its promoter in 193H and 232H, and negative MP-divergence in 150H (albeit not significant) (SI table 3; Fig. 5C). This candidate was the only one of the five genes in the FT QTLs^{epi} for which also a substantial effect on RNA expression was observed: the gene was significantly (> 2 fold from MPV) downregulated in 92H with methylation levels around MPV, while it was expressed around mid-parent levels in 150H and 193H and slightly downregulated in 232H (Fig. 5C, SI table 4). These results reveal that the relation between differences in DNA methylation at the promoter region and RNA expression is not clear. Possibly, DNA methylation profiles at regulatory elements not located in the promoter regions play a role here. Nonetheless, expression levels of this gene could still play a role in our heterosis phenotypes: the transducin gene was significantly downregulated in 92H, which was the only one of the four epiHybrids that did not show leaf area (LA) heterosis (Chapter 3). In the other epiHybrids the gene was expressed around mid-parent levels (with some deviations in both directions) whilst H150 displayed positive mid-parent heterosis (MPH), and H193 and H232 Best-parent heterosis (BPH) for LA (Chapter 3). While nothing is known on the function of this specific protein, WD proteins are often related to developmental processes in Arabidopsis [129,150]. This Transducin/WD-40 gene has partially conserved protein domains with Gigantus1 (GTS1), which controls seed germination, biomass accumulation and ribosome biogenesis in Arabidopsis [129,150]. Should AT3G26480 function in similar processes a relation with leaf area heterosis is possible.

AT3G61470 encodes a component of the light harvesting antenna complex of photosystem I. The gene showed a significant increase in gene-body methylation in 92H (SI table 3; Fig. 5D) and expression levels were around

MPV-level in all epiHybrids. There was slight positive divergence for 92, 150 and 193H (strongest for the last).

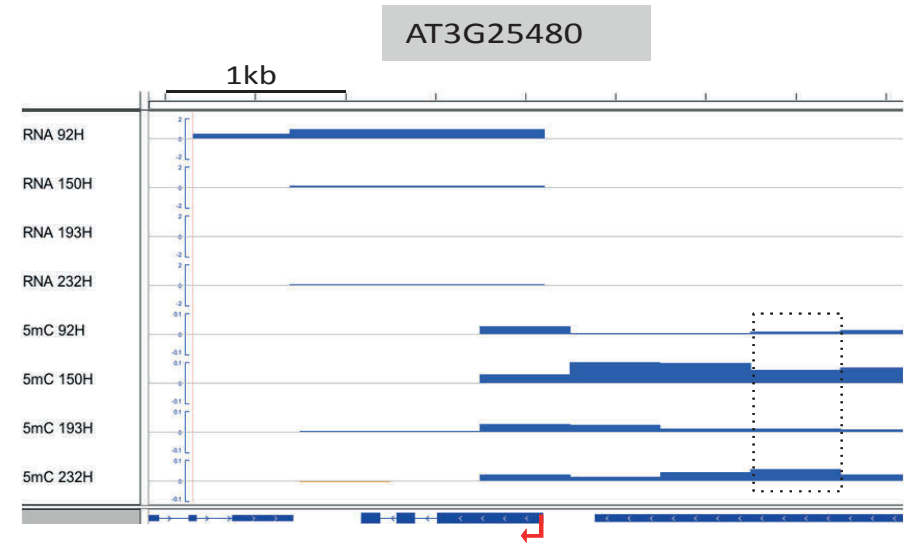
AT3G61480 displayed significantly increased promoter methylation in 92H, which was associated with (non-significant) below MPV gene expression (SI table 3; Fig. 5E). This may be a case were DNA methylation at the promoter affects gene expression in 92H. Methylation and expression levels in the other epiHybrids were around MPV (slightly above MPV for 150H and 232H) and at or close to MPV-level. The gene encodes a Quinoprotein amine dehydrogenase, an enzyme with unknown function.

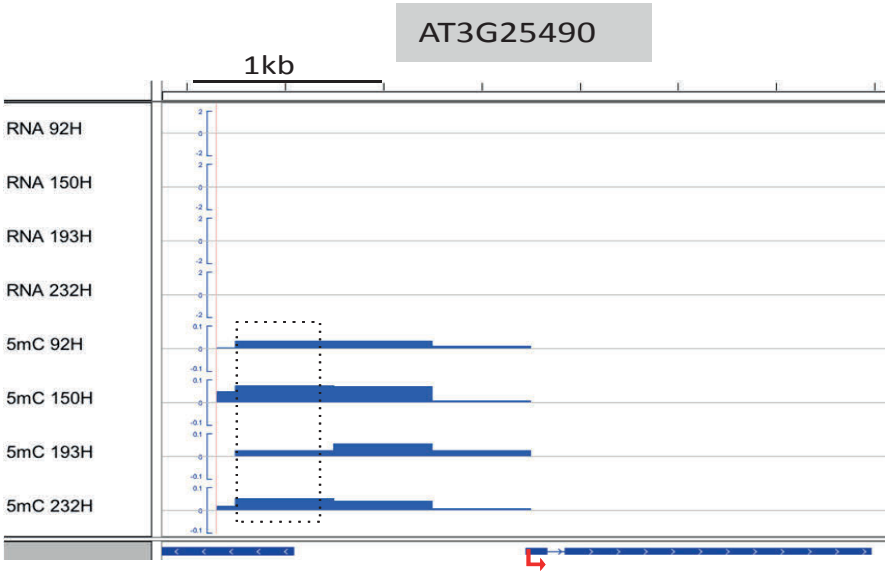
In sum, significant non-additive DNA methylation divergence in promoter or gene body regions of genes could not be clearly linked to significant non-additive gene expression changes in these five selected genes. Also, none of the other genes that were selected on the basis of DNA methylation divergence (SI table 3), showed significant non-additive gene expression.

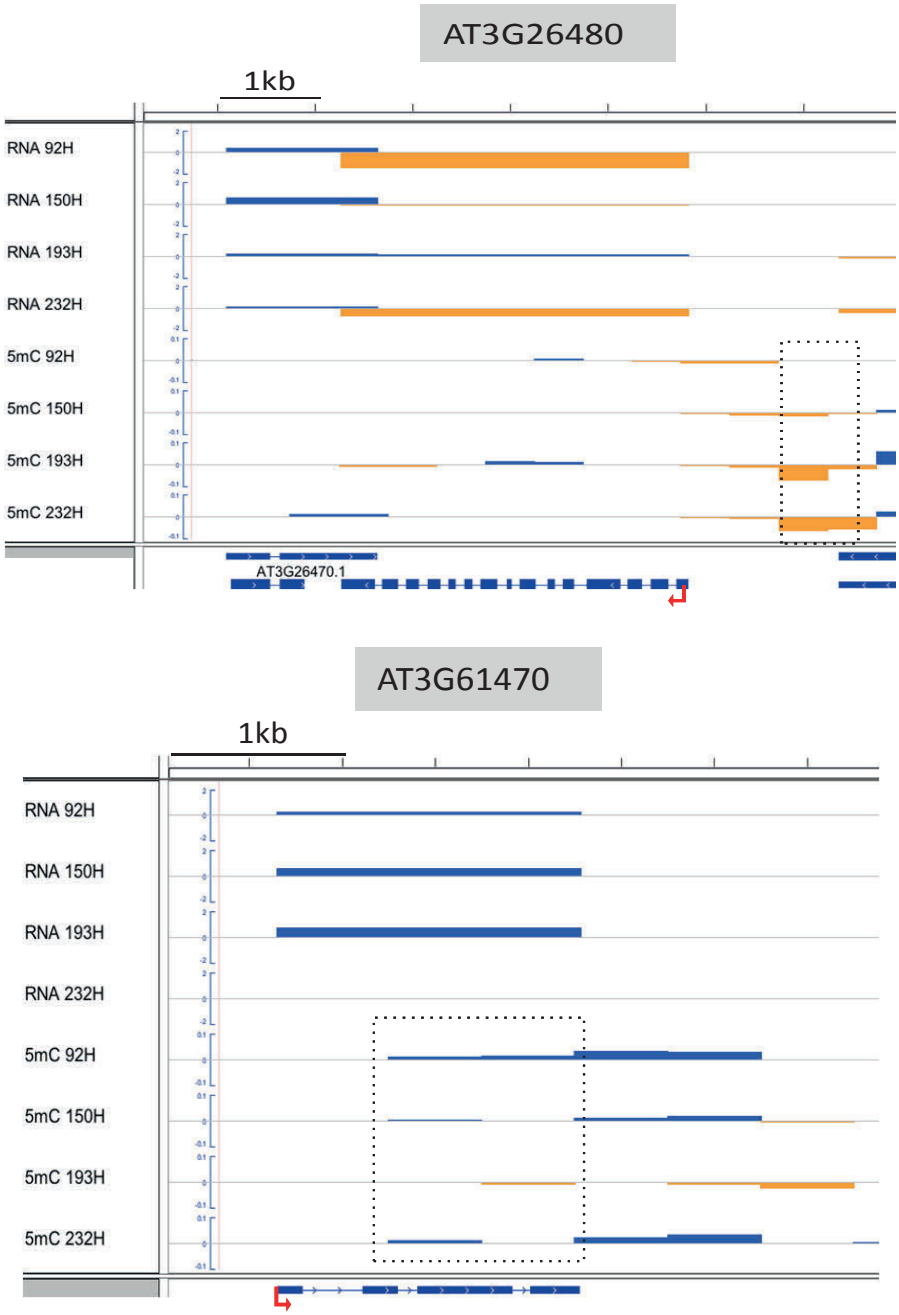
In our second approach, we identified candidate genes in the $QTLs^{epi}$ regions based on significant RNA expression changes from MPV (> 2 fold difference from MPV). We found 8, 1, 3 and 6 genes up- and 12, 6, 2, and 9 genes downregulated in epiHybrids 92H, 150H, 193H and 232H, respectively (SI table 4). Several of those up- or downregulated genes were shared among the epiHybrids. We shall below discuss just one example of a gene in this selection.

In the HT_QTL^{epi}, a gene involved in oxidative stress response (AT4G21840) [151] was significantly downregulated in 92H, 150H and 232H (Fig. 6D, SI Fig. 8A, SI table 4), while it was expressed close to MPV level in epiHybrid 193H. In a prior study in Arabidopsis hybrids, AT4G21840 was also downregulated (in C24 x Ler and C24 x Col-wt hybrids) displaying heterosis for rosette diameter [152]. EpiHybrid 193H showed best-parent heterosis (BPH) in plant height and 150H showed positive MPH, while 92H and 232H did not display HT heterosis. We wonder if having expression levels close

to MPV could be linked to BPH in plant height (Fig. 6D). Pathways can be dosage sensitive for the amount of particular gene products, hence very specific gene expression levels can be relevant to mediate heterosis [153]. While we describe here mostly candidate genes that were selected based on significant MP-divergence in DNA methylation and/or RNA expression, additive expression of genes located in the QTL^{epi} regions (or network of genes) may play a role as well in our observed phenotypes.


AT3G25480
1kb
RNA 92H
RNA 150H
RNA 193H
RNA 232H
5mC 92H
5mC 150H
5mC 193H
5mC 232H


AT3G25490
1kb
RNA 92H
RNA 150H
RNA 193H
RNA 232H
5mC 92H
5mC 150H
5mC 193H
5mC 232H


AT3G26480
1kb
RNA 92H
RNA 150H
RNA 193H
RNA 232H
5mC 92H
5mC 150H
5mC 193H
5mC 232H
AT3G26470.1
AT3G61470
1kb
RNA 92H
RNA 150H
RNA 193H
RNA 232H
5mC 92H
5mC 150H
5mC 193H
5mC 232H

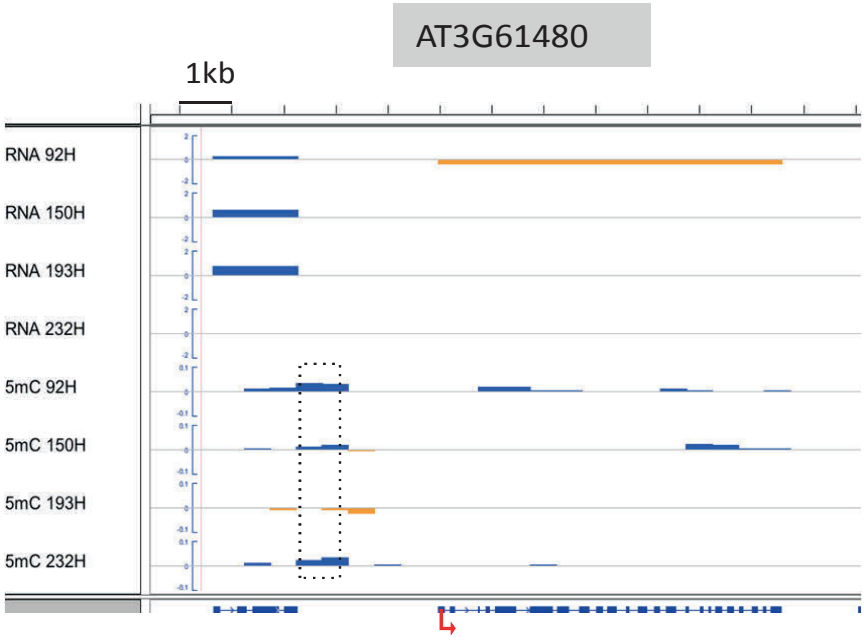


Figure 5: Genome-browser view of five candidate genes from the QTL^{epi} regions selected based on differential DNA methylation levels (for criteria see Methods and SI table 3). Methylation difference from MPV (range -0.1 to 0.1) and log2 fold RNA expression divergence from MPV (range -2, 2) is depicted. Positive divergence is shown in blue, negative divergence in orange. The transcription direction of the candidate gene is indicated with a red arrow. The areas of interest are indicated with dashed boxes.

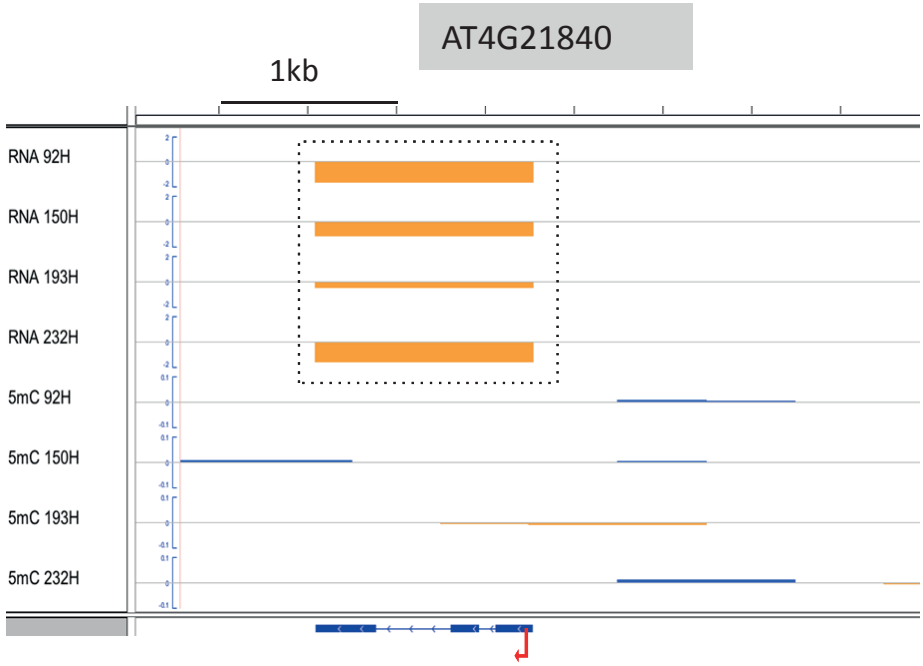


Figure 6: Genome-browser view of one candidate gene from the QTL^{epi} HT_1 selected based on RNA expression divergence (SI table 4). Methylation divergence from MPV (range -0.1 to 0.1) and log2 fold RNA expression divergence from MPV (range -2, 2) is depicted. Positive divergence is shown in blue, negative divergence in orange. The transcription direction of the candidate gene is indicated with a red arrow. The areas of interest are indicated with a dashed box.

Discussion

Previously, we created epigenetic hybrids (epiHybrids) by crossing Col-wt with several different epiRILs and described epigenetic divergence as being sufficient for heterotic phenotypes of those hybrids (Chapter 3). In addition, we associated our heterotic phenotypes in flowering time, leaf area and plant height with a number of QTLs^{epi}. To understand the molecular basis of these heterotic effects, we here analyzed the transcriptomes and methylomes of four selected epiHybrids and their

parental lines in more detail. Our analysis uncovered widespread non-additive transcript level changes, mostly in transposons, but also in protein coding genes, and provided indications for global methylome interactions that could be the result of TC(d)M events. Analyses of the entire genome as well as of the identified QTL^{epi} regions revealed non-additive changes in epiHybrids both in terms of DNA methylation as well as in terms of RNA expression. We hypothesize that such changes may constitute molecular mechanisms underlying the heterotic phenotypes, possibly by DNA methylation altering the expression states of phenotypically relevant genes. Those probably include genes within $QTLs^{epi}$ we identified in Chapter 3. Which expression states would be causal remained elusive however.

Trans-chromosomal (de)methylation events in epiHybrids

Trans-chromosomal (de)methylation (TC(d)M) events are shown to occur in genetic as well as in a *met1*-derived epigenetic hybrid [78,79,146]. They refer to *in trans* communication between differentially DNA methylated regions resulting in homologous sequence regions acquiring a similar methylation profile. This can result in gain but also loss of DNA methylation. Small RNAs have been suggested to mediate such methylome interactions [78,79,146] through a pathway termed RNA-dependent DNA methylation (RdDM). In fact, RdDM has recently been demonstrated to be essential for TC(d)M events in Arabidopsis hybrids [146]. However, not all TC(d)M events could be linked with the occurrence of small RNAs, indicating that other factors may also play a role in TC(d)M [79,146].

We observed indications for TC(d)M events in all four epiHybrids, with loss of DNA methylation occurring in 0.5 to 1.3% of all genomic windows (sliding window approach using a size 1kb and steps size of 500bp) and gain of methylation in 1.2 to 2% of all windows . While this could point towards methylation gain occurring more often than methylation loss, it

could also be the result of subtle biases related to PCR amplification during sequencing library construction [147]. Most TCdM events occurred at regions that were differentially methylated in the parents, which is in line with a prior study on a genetic Arabidopsis hybrid between the accessions C24 and Ler [79]. This indicates that a large fraction of TCdM interactions are allelic. On the other hand, the majority of TCM events appeared to occur at regions that did not show significant methylation differences between the parents, consistent with a study on another Arabidopsis hybrid (hybrid between Col and C24) [146], pointing towards methylome interactions happening at regions with similar methylation profiles and/or a role of non-allelic interactions. We did not investigate the locations of our TC(d)M events in detail, however it appeared that they occur at both epi-homozygous (Col-wt/Col-wt) and epi-heterozygous (*ddm1*/Col-wt) regions in the epiHybrids (Fig. 3A).

We also looked into whether the putative TC(d)M regions were shared among the four epiHybrids. We found only few shared TCdM and TCM events. The largest fractions of TCM events were unique to the individual epiHybrids (Fig. 3B). This could be interpreted as the specific DNA methylomes of the different epiRIL parents strongly influencing methylome interactions in the epiHybrids. To understand the contribution of *ddm1*-derived regions to the occurrence of TC(d)M events better, their contribution to TC(d)M needs to be analyzed in more detail, for instance, by investigating TC(d)M at genomic regions that share particular DNA methylation profiles among the epiRILs.

So, do such methylome interactions play a role in causing heterosis of phenotypic traits? A recent report has indicated that eliminating the RdDM pathway in Arabidopsis does not disrupt heterosis in leaf rosette radius, at least not in 20 day old seedlings [146]. Similar observations have been made in maize hybrids (B73 x Mo17), where despite a mutation in the RdDM pathway and corresponding reduction in small RNAs, heterosis

in cob weight was still observed [45]. This implies that TC(d)M events are not crucial for heterosis, at least not in the two instances studied. In an epiHybrid generated from a *met1* mutant and Col-wt, thousands of TC(d)M regions have been identified, but the resulting hybrid was not heterotic for the trait measured (biomass) [78]. These results indicate that TC(d)M events per se is not sufficient to cause heterosis, at least not in the two instances studied.

TC(d)M events that result in a heritable change of the DNA methylation profile would fall under the definition of paramutation. The inheritance of DNA methylation patterns acquired by TCM in Arabidopsis hybrids has been investigated at some DNA sequence regions and these were (mostly) stably inherited to the F2 generation, indicating that these regions, at least partially, show paramutation-like behavior [121,146]. Interestingly, one example of a stably inherited TCM event in an Arabidopsis hybrid was not detected in 15 DAS seedlings, but was observed in floral buds, which are formed only quite late in F1 development [121]. This indicates that the developmental time point studied may have an influence on which TC(d)M events are detected. Perhaps, TC(d)M events related to certain phenotypic traits can only be detected later in development due to accumulation of methylation over time.

While hampering TC(d)M at some sites may not disrupt heterosis of certain traits, TC(d)M events at other genomic locations may still play a role in heterosis in other traits, or in hybrids between accessions or species different from those currently tested. Future research will need to investigate how this genome-wide occurrence of TC(d)M affects gene expression and relates to phenotypic variation.

Transcriptional activation of transposons in the epiHybrids

As compared to Col-wt, n the epiRIL parental lines and their respective epiHybrids, transposable elements (TEs) were transcriptionally activated.

In the epiHybrids the transposon families with the highest number of transposons showing significant divergence in RNA expression from mid-parent value (MPV) belonged to the LTR/*Gypsy* elements, DNA/*MuDR* and RC/Helitrons families (table 2). For other TE families, including Line/L1 and LTR/*Copia* elements, we also found cases with significant divergence in RNA expression from MPV, these occurred however not as frequent as in the aforementioned families.

The TE families affected in the epiHybrids belonged to TE classes having different transposition mechanisms and preferential genomic locations [149]. For instance, LTR/*Gypsy*, LTR/*Copia* and LINE elements are retrotransposons (Class I), meaning that they propagate via an RNA intermediate. *Gypsy* elements are concentrated in pericentromeric regions, while LTR/*Copia* and LINEs are more dispersed throughout the genome [18,149]. *MuDRs* are DNA transposons (Class II) that are evenly dispersed throughout the Arabidopsis genome [18,148,149]. Most DNA transposons transpose via a "cut and paste" mechanism without an RNA intermediate [149]. The RC/Helitrons constitute a special type of DNA-transposon, which are mostly located in euchromatin and presumably propagate via a rolling-circle mechanism, whereby one strand of the TE is transferred to a new genomic location and serves as the template for DNA synthesis [148,149].

A previous study reported that a *ddm1* mutation affects DNA methylation mainly at large transposons [18]. Since the epiRILs used to create the epiHybrids were derived from a *ddm1* mutant [52], we investigated whether there is an effect on transcript levels of predominantly long TEs. Indeed, we found the average sizes of up- and downregulated TEs in the epiHybrids were larger than the average sizes of TEs in general.

Transcriptional activation of transposons usually does not coincide with their mobilization [126]. TE transcripts may partially induce a re-silencing mechanism of TEs that lost DNA methylation, as has been previously

described in *ddm1*-mutant plants [154]. In a *ddm1* mutant plant, an increase in transcripts has been reported for 2,294 TEs [18]. However, one report documented mobilization of only a few TEs (mostly *Copia* and *Gypsy*), and only after several generations of inbreeding [126]. Nonetheless, TE transposition has been described in some *ddm1*- and *met1*-derived epiRIL lines [20,138], lines in which the mutations are not present anymore [52,53].

Hence, TE mobilization may occur to a certain extent in our epiRIL parental lines (and epiHybrids) and might affect gene expression and consequently our heterotic phenotypes. (Novel) TE insertions can have various effects in genomes, and genes in particular [155]. Besides the fact that TEs can cause severe detrimental effects by knocking out genes, they can also induce gene expression changes by other means [155,156]. Such changes can be mediated, for instance, by spreading of DNA methylation from TEs into flanking genes, by TE sequences acting as regulatory sequences or by TE-derived small RNAs regulating the expression of a host gene [155,156]. TEs can even be involved in the creation of "novel" gene variants [155]. For example, helitrons in maize have been found to be very potent in capturing sequences from multiple genes during their transposition [155]. This can result in the creation of "chimeric" genes inside TEs, whose transcripts can produce numerous final products [155]. In addition, fragments of TEs have been found within coding regions of genes, with estimates of > 1 % of known protein-coding genes in Arabidopsis harboring some TE-related segments [157].

Arabidopsis has a rather small, gene-dense genome, consisting of around 36% of TE sequences, which are mostly localized in the pericentromeric regions [30,31]. TE mobilizations could have strong implications for gene expression and consequently phenotypes. For the epiHybrids, it is possible that hybridization resulted in the Col-wt genome compensating for strong, TE-caused detrimental effects originating from the epiRIL parents'

genomes. This may have contributed to the heterotic phenotypes observed. To understand if TE insertions do contribute to gene expression changes and potentially the heterotic phenotypes in the epiHybrids, novel TE insertions would have to be monitored and associated with gene expression levels.

Gene expression patterns potentially relevant for heterosis

Highly specific expression levels might play a role in heterosis. This does not only concern strong non-additive changes in expression, but it could also involve gene expression levels around mid-parent value. It all depends on the optimal dosage of particular gene products.

A good example of a heterotic dosage effect where functional trait depends on expression level of a gene, has been described for a locus in tomato: the SINGLE FLOWER TRUSS (SFT) gene, which in its heterozygous (overdominant) state resulted in yield increases as compared to both the homozygous wildtype and the mutant parental lines [158]. SFT encodes the flowering hormone florigen and heterosis is linked to dosage sensitivity of the florigen pathway, whereby the expression level of SFT in the hybrid lies between that of the two parents [159]. Similarly, it has been suggested that an intermediate expression level of SA-related genes in between the parental levels was optimal, and would mediate seedling growth heterosis in Arabidopsis hybrids derived from C24 and Col-wt [153]. Those conclusions were drawn from finding that in *ddm1* mutant background heterosis is reduced and that correlates with increased expression of SA-related genes [153].

In the present study, a complex example of such mRNA-to-trait heterosis arising in the pathway between mRNA and trait, is the oxidative stress response gene located in the HT QTL^{epi}. This gene showed significant negative non-additive expression in 92H, 150H and 232H - but displayed an expression level closer to mid-parent value in 193H. Yet, the epiHybrid

193H displayed best-parent heterosis (BPH) for plant height, while neither 150, 92H nor 232H exhibited heterosis.

In this study, we discussed genes that were selected based on significant non-additive divergence in DNA methylation and/or RNA expression. However, additive expression of candidate genes (or combinations of candidate genes) from the QTL^{epi} regions may just as likely underlie our heterotic phenotypes.

Preliminary insights into genes at the QTL^{epi} regions

In Chapter 3, we created and phenotyped nineteen epiHybrid lines and detected two such $QTLs^{epi}$ for flowering time (FT) and one for plant height (HT) heterosis, located on chromosomes 3 and 4, respectively. We found that the same two regions identified as putative FT $QTLs^{epi}$ were relevant for leaf area (LA) heterosis (same regions detected during mapping but below significance threshold). All three QTLs were associated with differentially methylated regions in the parental genomes.

In the present study we went further, assuming that each QTL^{epi} could be due to expression states at single or few genes. We investigated non-additive DNA methylation and gene expression profiles in detail for genes within the QTL^{epi} regions that we selected on the basis of a substantial number of criteria. The approach based on DNA methylation data faced limitations, such as the definition of promoters as 1.5kb regions upstream of the transcription start site. While such approximations may be legitimate in genome-wide datasets, they may be suboptimal here. If promoter regions are not correctly identified, associating changes in their methylation level with expression changes is severely hampered. Moreover, some genes may not be expressed despite significant methylation differences being detected between the lines resulting in the identification of candidate genes that are neither expressed in parental nor in epiHybrid lines (as seen for AT3G25490).

For the selection based on RNA expression, we simply selected candidates showing significant (>2fold from mid-parent value) non-additive expression changes and highlighting candidates that were shared in more epiHybrids.

Our preliminary attempt to identify DNA methylation or expression at genes that strongly correlated with the heterotic events, did not lead to a clear outcome. This is not to say that new attempts using different (re-evaluated) criteria might not be more successful in the future. In the end, we believe it is a more fruitful approach to first identify interesting gene expression patterns, followed by investigating potential differences in DNA methylation at candidates of interest (including careful determination of promoter regions). Because ultimately, mainly DNA methylation that has an effect on gene expression will be relevant for phenotypic traits. A different viewpoint is of course that phenotypic traits are complex functions of networks and it may well be unrealistic to assume that expression of single factors or genes to will always be responsible. The very fact that we did find $QTLs^{epi}$ remains statistically sound and deciphering the effects of those regions will be a goal of future research efforts.

Multilevel Heterosis

In this study we have observed heterosis at multiple levels. First some methylation levels of epiHybrids were not additive - this was exemplified by strong indications for trans-chromosomal (de) methylation (TC(d)M) events. Second, we observed non-additive gene expression in the epiHybrids globally and within previously identified QTL^{epi} regions. And of course, it all started with heterosis in terms of phenotypic traits. The simplest possible mechanism would be one where TC(d)M-like events would change methylation profiles (at promoters) which would affect

transcription levels and in turn cause a heterotic effect in the quantitative trait (proportional to the transcription level). In our preliminary analyses this is not what we observed however: in the QTL^{epi} regions, we found non-additivity of mRNA levels for genes that did not show significant non-additivity in terms of methylation and vice versa. And further we could not identify single key genes completely linked to the quantitative traits, suggesting that the identification of key genes in such way is not that simple either.

Material and Methods

Plant Material

As described in Chapter 3.

DNA Sample preparation for whole-genome Bisulfite-sequencing

Plant material for sequencing was grown under the same controlled environmental conditions as described above. Aerial rosette tissue at 21/22 DAS (days after sowing) was harvested before noon at two consecutive days and snap-frozen immediately in liquid nitrogen. Material was stored at –80°C until processing.

Genomic DNA from two biological replicates (2 x 6 rosettes) was extracted using a standard CTAB-based extraction protocol followed by an RNase digest. 5µg DNA per sample were submitted to the Beijing Genome Institute (BGI) for Bisulfite treatment, library construction and sequencing. Sequencing (whole-genome bisulfite sequencing; WGBS-seq) was performed on an Illumina HiSeq4000 instrument. Reads were sequenced up to 150 bp (paired-end) for libraries with an insert size of 200 bp.

Alignment of bisulfite treated reads and construction of DNA methylomes

Read sequences were quality trimmed and adapter sequences were removed with the use of Cutadapt (version 1.9; python version 2.7.9; [160]). Trimming was performed using the paired-end mode and the quality threshold was set to a phred score of 20 (q = 20). We applied the default error rate of 10% for the removal of the adapter sequences. Afterwards we discarded reads that were shorter than 40 base pairs.

Reads were subsequently mapped to an index genome with the use of BS-Seeker2 (v2.0.10; [161]). The maximum allowed proportion of mismatches was set to 0.05 (m = 0.05) and the maximum insert size was set to 1000 bp (X = 1000). Bowtie2 (version 2.2.2; [162]) was used for the alignment of the reads. Mapping was done in two steps. Read pairs were first aligned in paired-end mode. Reads that could not be aligned because they could not be aligned concordantly were subsequently aligned in single-end mode as described on the website of BS-seeker2 (http://pellegrini.mcdb.ucla.edu/BS_Seeker2/). Mate 2 reads were first reverse complemented before alignment.

Duplicate read pairs were removed in paired mode using in-house scripts by comparing start positions (5' ends) of the mates and the orientation (mapped to forward or reverse strand). Samtools (version 1.2; using htslib 1.2.1; [163]) was used for sorting the mapped read pairs in this procedure. When only one mate of the pair was aligned duplicate reads were removed by comparing only with read pairs for which only one mate was aligned (same strategy; comparing start of the alignment and orientation). In both cases (paired and single mode) we kept one read pair (randomly chosen). Also when both mates of one pair were overlapping each other we corrected for the overlap by trimming the reads until the middle of the overlapping part. When one mate contained a mismatch at a cytosine position and the other mate contained a methylation call (methylated or unmethylated) we kept the call even if it came from the mate for which this overlapping part was discarded (or trimmed).

After removal of duplicate read pairs the sam files were sorted using Samtools [163] and methylomes were subsequently reconstructed. With the use of the information in the sam files that indicates which reference cytosines are unmethylated (a T mapped to a C; bisulfite conversion) or methylated (a C mapped to a C; no conversion), the number of reads indicating methylation and the total number of reads were determined for each reference cytosine. After determination of these frequencies the methylation level of each cytosine was determined. This methylation level is defined as the proportion of reads that indicate that the cytosine is methylated (proportion of non-converted cytosines).

Identification of DNA methylation outlier windows

After the construction of the methylomes, average methylation levels were calculated for genomic windows with a size of 1 kb (step size 500 bp; sliding windows). For this analysis only cytosines with a coverage of three or more reads were selected (sufficient read coverage). Average methylation levels were calculated for all cytosines and for each context separately (CG, CHG and CHH methylation levels). In order to focus on robust methylation levels, we required a minimum number of cytosines with sufficient read coverage for each window. The minimum for the calculation of the methylation level based on all cytosines and those of the CHH context was set to 50 cytosines; those of CG and CHG was set to 10 cytosines. Analysis was based on 93.0 till 97.9 % of all genomic windows. Methylation levels, which were based on a lower number of cytosines were discarded. The mid-parent methylation level was calculated as the average of both parents (average methylation level wild-type and epiRIL). Mid-parent divergence was calculated by subtracting the mid-parent methylation level from the average methylation level of the F1 hybrid. Outlier windows were defined as windows for which the mid-parent divergence was higher than three SDs (SD: standard deviation) from zero

or lower than minus three SDs from zero (zero: no difference; see Figure Sx).

Selection candidate windows in the QTLs^{epi} regions based on DNA methylation

Several filters were applied for the selection of potential candidate windows. Outlier windows were considered candidate windows when they met the following criteria:

Outlier windows should be located within one of the QTL^{epi} regions and have an overlap with the promoter or the body of genes. Genes as defined by TAIR (TAIR 10 release). Promoter region is defined as the 1.5 kb region upstream from the TSS (Transcription Start Site).

Outlier windows should have a match with significant difference in methylation levels between the two parents. For the analysis of all cytosines and cytosines with CHH context the difference in methylation level (epiRIL minus wild-type) should be lower or equal to -0.1 or higher or equal to 0.1. For the analysis of the cytosines with CG and CHG context the difference in methylation level should be lower or equal to -0.2 or higher or equal to 0.2. The methylation difference between the parents should be in the same direction for trios with methylation divergence (divergence from mid-parent value) in the same direction.

Outlier windows should match with the observed haplotypes. All trios with either wild-type or *ddm1-2* haplotype should have an outlier window. The direction of the mid-parent methylation divergence should be in the same direction (for all wild-type trios the mid-parent methylation divergence should be in the same direction or for all *ddm1-2* trios the mid-parent methylation divergence should be in the same direction). When all trios have an outlier window the direction should be opposite for wild-type and *ddm1-2* haplotypes.

Mix of directions and borderline cases: In case an outlier window has different directions in different hybrids one of the directions should match with significant differences in methylation levels between parents and match with one of the haplotypes. A few borderline cases (borderline significant) were also included as well as cases where the direction for mid-parent divergence and/or difference in methylation levels between parents matches with the haplotypes.

RNA sample preparation for RNA-sequencing

Plant material for sequencing was grown under the same controlled environmental conditions as described above. Aerial rosette tissue at 21/22 DAS (days after sowing) was harvested before noon at two consecutive days and snap-frozen immediately in liquid nitrogen. Material was stored at –80°C until processing.

RNA was prepared for two or three biological replicates (2/3 x 5 rosettes). Two biological replicates were prepared for lines: 92, 92H, 232. Three biological replicates were prepared for lines: 150, 150H, 193, 193H, 232H, Col-wt.

RNA was extracted using a protocol combining Trizol-based (TRI reagent, Sigma, T9424) extraction with the RNeasy Plant Mini Kit (Qiagen, Cat No./ID: 74903). Frozen plant material was ground in liquid nitrogen and TRI reagent was added to the frozen powder (~1ml per 100mg tissue). After an incubation of 5-30 minutes at room temperature chloroform/isoamyl alcohol (ratio 24:1) was added, mixed and incubated for another 3 minutes at room temperature. Next, the samples were centrifuged for 15 minutes at 12 000 x g at 4°C and the upper phase was transferred into a new tube. 1 volume isopropanol was added and samples were incubated at room temperature for 10 minutes before transferring them to an RNeasy MINI spin column (supplied in the RNeasy Plant Mini Kit). A centrifugation step for 15 seconds at 9 000 x g was performed, flow-

through was discarded and 700 µl buffer RW1 (supplied in the RNeasy Plant Mini Kit) was added to the column followed by another centrifugation step for 15 seconds at 9 000 x g. 500 µl buffer RPE (supplied in the RNeasy Plant Mini Kit) was added to the column and sample was centrifuged again for 15 seconds at 9 000 x g. Again, 500 µl buffer RPE was added to the column and sample was centrifuged for 2 minutes at 9 000 x g. Next, the column was transferred in a new 2ml collection tube followed by a centrifugation step for 1 minute at high speed to dry the membrane. Again, the column was transferred in a new 1.5ml collection tube and 50ul RNase-free water was added directly on the column membrane. The RNA was eluted by centrifugation for 1 minute at 9 000 x g.
To remove traces of DNA, a DNase I digest was performed (DNA-free DNA removal kit, Ambion, AM1906) according manufacturers instructions.
1-2µg RNA were submitted for library preparation and sequencing on an Illumina HiSeq2500 instrument to Wageningen University (WUR). 50bp single-read or 125bp paired-end read sequencing was performed.

RNA-sequencing

The 125 bp paired-end reads were trimmed to 50 bp using Fastx toolkit [164] and only one end was used for read mapping to make the analysis comparable to the other libraries with 50 bp single-end reads. The sequenced reads were trimmed at the both ends based on sequencing quality (Q20) and remaining Illumina adaptor sequences were removed using Trimmomatic [165]. When the remaining read length was less than 35 bps, the read was removed from the analysis. The reads were aligned, allowing 1 mismatch, to the reference genome, Arabidopsis thaliana TAIR10 assembly obtained from Ensembl Plants [166], using TopHat2 [167] and Bowtie [168]. The transcript assembly and gene expression level calculation for each replicate were performed with a guided reference (TAIR10, [166]) using Cufflinks pipeline (Cufflink, Cuffquant and Cuffnorm;

[169]). Only uniquely-mapping reads were retained for further analysis. To determine the reproducibility, the RPKM values for each gene in every pair of replicates were plotted against each other and Spearman's rank correlations were calculated in R [94]. The RPKM values for each sample were calculated with multi-read correction using Cuffdiff [93]. The log-fold changes between the F1 and the parental epiRIL and between F1 and mid-parents value were calculated and log transformed (log base 2).

Acknowledgments

We thank H. Westerhoff for critically reading the manuscript.

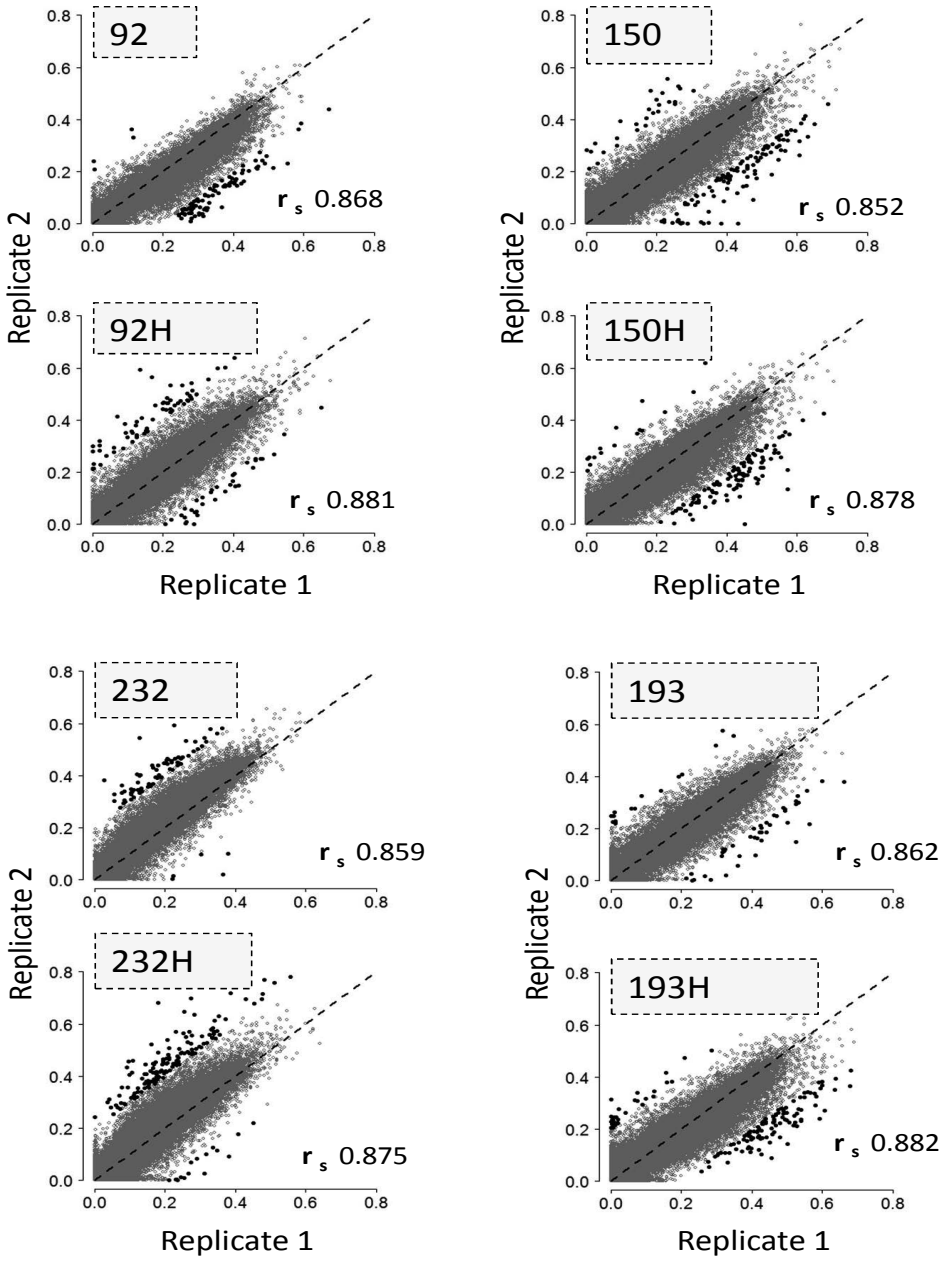
Funding

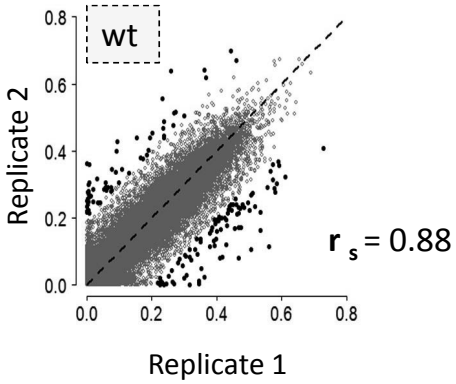
K.L. was supported by the Centre for Improving Plant Yield (CIPY) (part of the Netherlands Genomics Initiative and the Netherlands Organization for Scientific Research). R.O. was supported by the European Commission 7th Framework-People-2012-ITN Project EpiTRAITS (Epigenetic Regulation of Economically Important Plant Traits; 316965). D.T.A. was supported by the Dutch Technology Foundation STW, which is part of the 498 Netherlands Organization for Scientific Research (NWO), and which is partly funded by the Ministry 499 of Economic Affairs (OTP Grant 12385 to MS). F.J. acknowledges support from the Technical University of Munich Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Program under grant agreement #291763.

Author contributions

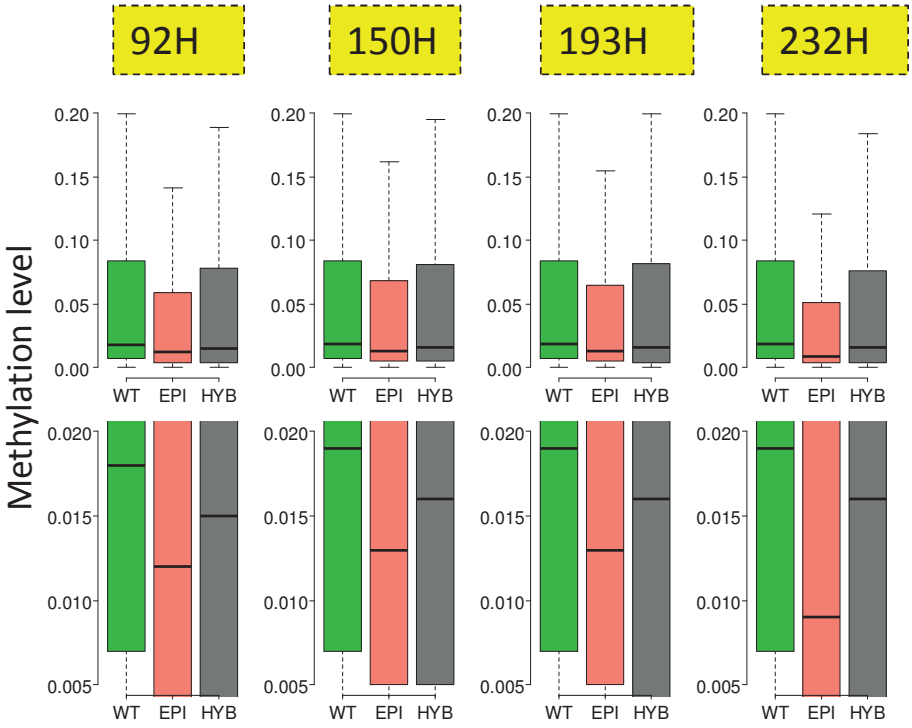
K.L, F.J and M.S planned and designed the study; K.L. did the experiments with help from D.A. and F.B; R.O. analyzed the RNA-seq. data; R.W. analyzed the Bisulfite-seq data; J.J.B provided material; K.L., R.W, F.J and M.S interpreted the data and K.L and M.S wrote the manuscript with contributions from F.J.;

Supporting Information

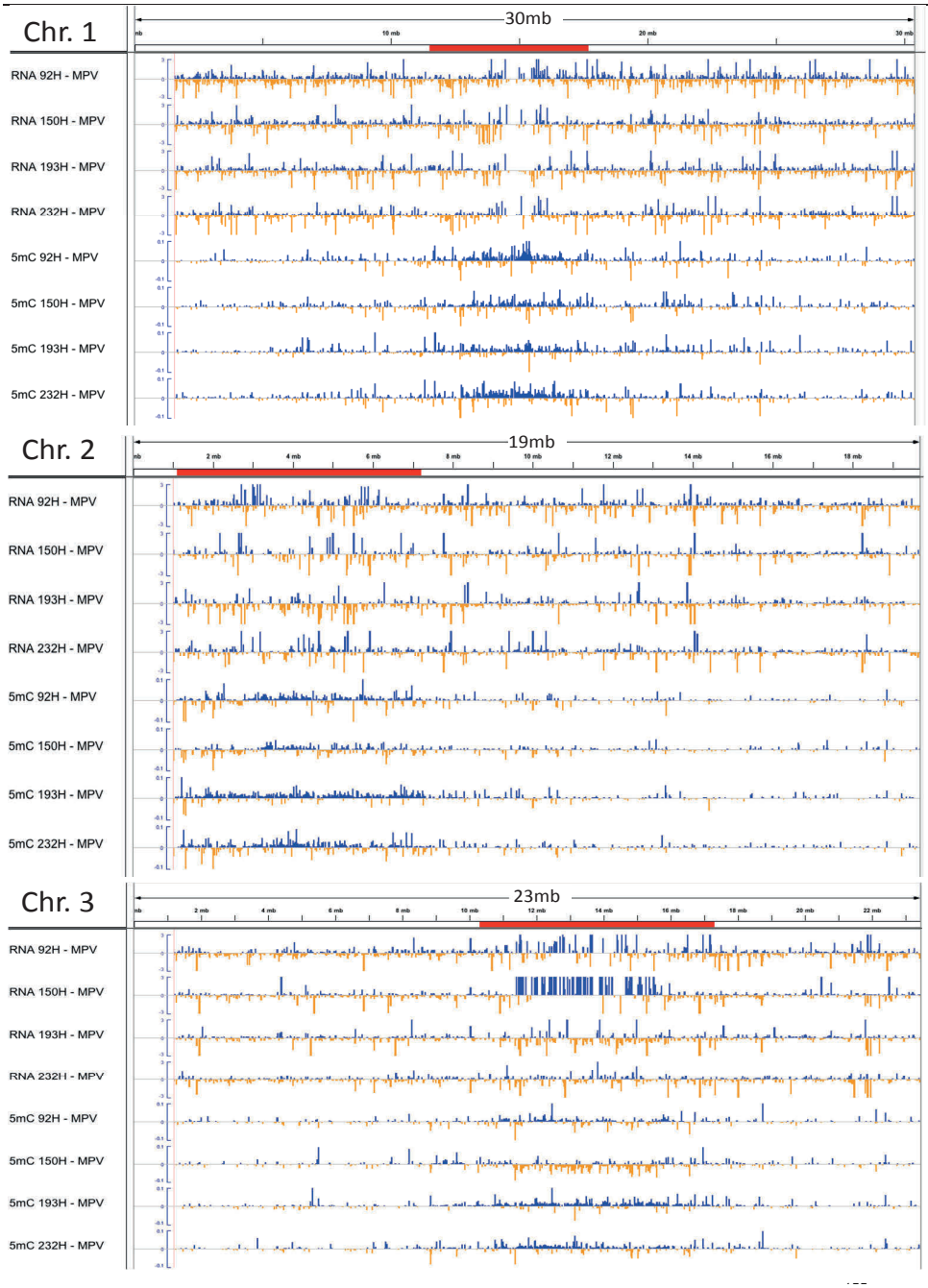


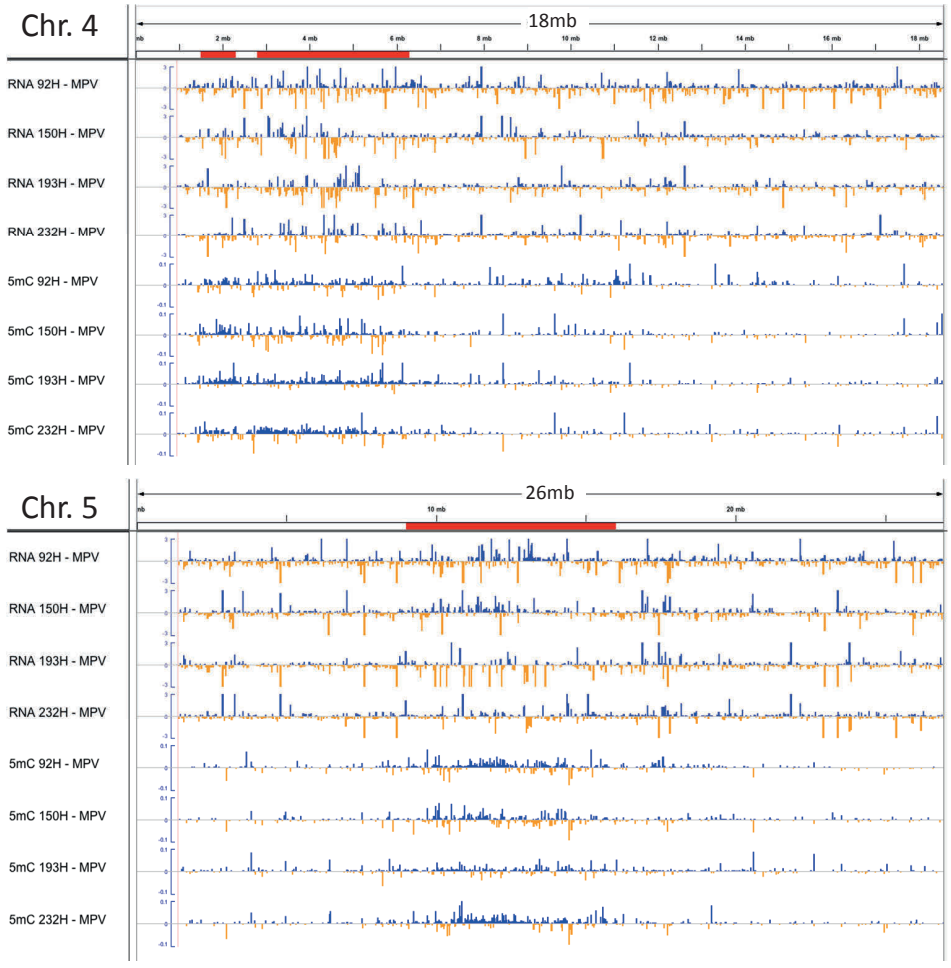


Supplemental figure 1: Spearman Correlation analysis between bisulfite-sequencing replicates. The correlation coefficient is denoted as r_s. Black colored windows indicate difference in methylation level of 0.2 (difference >= 0.2) or of -0.2 (difference <= -0.2).



Supplemental Figure 2: Global methylation levels in all epihybrids and their parental lines. Col-wt (green, WT), respective epiRIL-parent (salmon, EPI) and epihybrid (grey, HYB) are depicted. Lower panel shows a zoom into the upper panel to illustrate differences in the median. Sequencing was performed in two biological replicates for each sample and reads were pooled for analysis.


Chr. 1
30mb
RNA 92H - MPV
RNA 150H - MPV
RNA 193H - MPV
RNA 232H - MPV
5mC 92H - MPV
5mC 150H - MPV
5mC 193H - MPV
5mC 232H - MPV
Chr. 2
19mb
RNA 92H - MPV
RNA 150H - MPV
RNA 193H - MPV
RNA 232H - MPV
5mC 92H - MPV
5mC 150H - MPV
5mC 193H - MPV
5mC 232H - MPV
Chr. 3
23mb
RNA 92H - MPV
RNA 150H - MPV
RNA 193H - MPV
RNA 232H - MPV
5mC 92H - MPV
5mC 150H - MPV
5mC 193H - MPV
5mC 232H - MPV

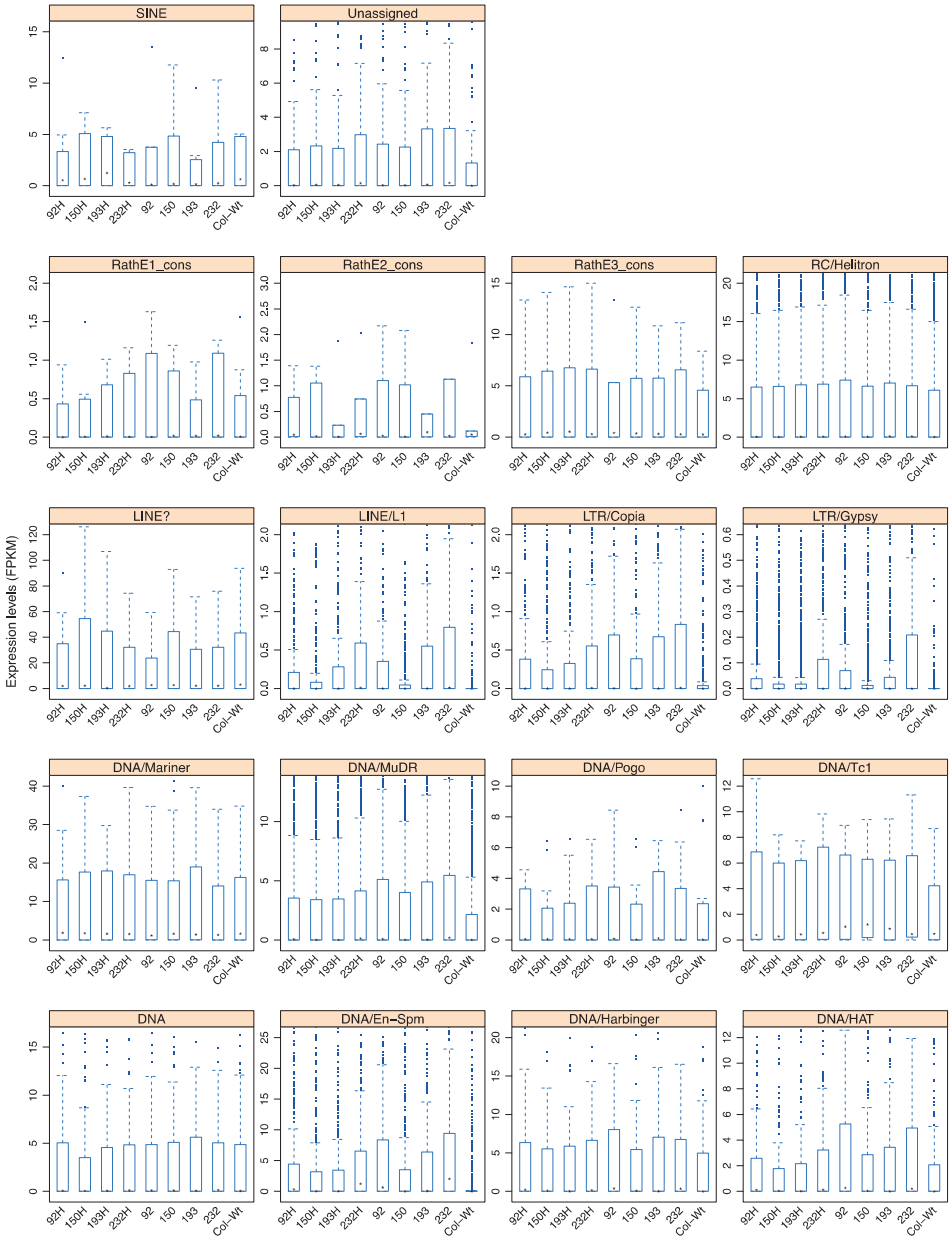


Supplemental Figure 2: Divergence from mid-parent values for RNA expression (RNA) and DNA methylation (5mC) in all epiHybrids. Positive divergence is shown in blue, negative divergence in orange. Depicted are difference for 5mC (range -0.1 to 0.1) and log2 fold difference for RNA expression (-3 to 3). The pericentromeres are indicated in red.

	92H_1	92H_2	150H_1	150H_2	150H_3	193H_1	193H_2	193H_3	232H_1	232H_2	232H_3	
92H_2	0,974											
150H_1	0,938	0,939										
150H_2	0,946	0,937	0,965									
150H_3	0,957	0,950	0,964	0,968								
193H_1	0,949	0,944	0,944	0,947	0,958							
193H_2	0,936	0,933	0,947	0,951	0,947	0,968						
193H_3	0,952	0,947	0,939	0,945	0,957	0,975	0,964					
232H_1	0,963	0,953	0,941	0,947	0,960	0,958	0,945	0,957				
232H_2	0,964	0,959	0,944	0,947	0,960	0,958	0,945	0,959	0,975			
232H_3	0,963	0,956	0,942	0,943	0,957	0,955	0,941	0,954	0,976	0,976		
92_1	0,970	0,964	0,928	0,933	0,952	0,944	0,929	0,944	0,962	0,957	0,959	
92_2	0,970	0,962	0,932	0,940	0,949	0,944	0,932	0,943	0,961	0,956	0,955	
150_1	0,952	0,946	0,961	0,966	0,967	0,943	0,942	0,937	0,946	0,947	0,944	
150_2	0,952	0,946	0,961	0,965	0,968	0,943	0,942	0,939	0,946	0,948	0,944	
150_3	0,959	0,952	0,949	0,958	0,972	0,947	0,933	0,946	0,956	0,955	0,953	
193_1	0,959	0,935	0,938	0,939	0,942	0,969	0,964	0,962	0,948	0,946	0,946	
193_2	0,936	0,939	0,924	0,927	0,943	0,969	0,951	0,964	0,950	0,950	0,948	
193_3	0,940	0,937	0,924	0,926	0,944	0,971	0,950	0,966	0,951	0,950	0,948	
232_1	0,957	0,953	0,929	0,932	0,946	0,943	0,928	0,942	0,968	0,970	0,972	
232_2	0,955	0,949	0,929	0,930	0,946	0,942	0,927	0,939	0,971	0,969	0,972	
ColWt_1	0,939	0,932	0,941	0,955	0,951	0,949	0,952	0,947	0,934	0,936	0,929	
ColWt_2	0,936	0,931	0,938	0,951	0,949	0,948	0,951	0,947	0,929	0,931	0,925	
ColWt_3	0,945	0,942	0,932	0,941	0,955	0,953	0,938	0,954	0,936	0,941	0,935	

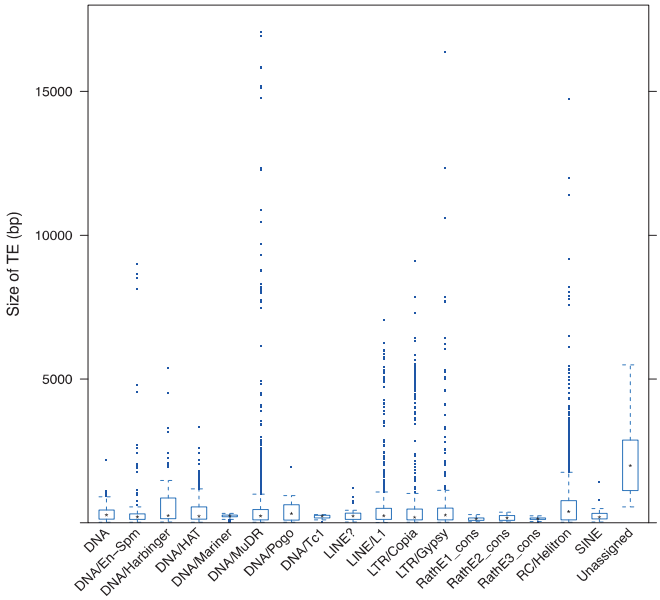
	92_1	92_2	150_1	150_2	150_3	193_1	193_2	193_3	232_1	232_2	ColWt_1	ColWt_2
92_2	0,973											
150_1	0,943	0,946										
150_2	0,942	0,945	0,975									
150_3	0,952	0,954	0,966	0,967								
193_1	0,932	0,935	0,935	0,934	0,934							
193_2	0,937	0,934	0,928	0,926	0,937	0,966						
193_3	0,938	0,934	0,928	0,926	0,938	0,968	0,978					
232_1	0,951	0,949	0,938	0,936	0,944	0,938	0,944	0,943				
232_2	0,951	0,950	0,937	0,936	0,945	0,938	0,946	0,944	0,976			
ColWt_1	0,927	0,929	0,949	0,947	0,941	0,941	0,930	0,930	0,914	0,912		
ColWt_2	0,924	0,926	0,946	0,944	0,939	0,937	0,928	0,928	0,908	0,905	0,978	
ColWt_3	0,934	0,931	0,942	0,940	0,949	0,935	0,940	0,941	0,920	0,918	0,966	0,970

Supplemental table 1: Spearman Correlation analysis between biological RNA-sequencing replicates. The correlation coefficient (r_s) is given for all combinations and replicates are highlighted.

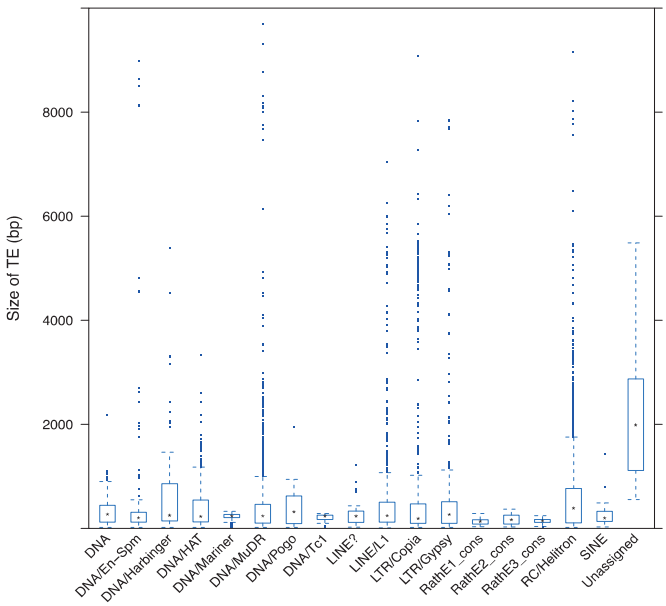


Supplemental Figure 3: RNA expression levels (uniquely-mapping reads) of transposon families in all lines. The black dot represents the median.

A

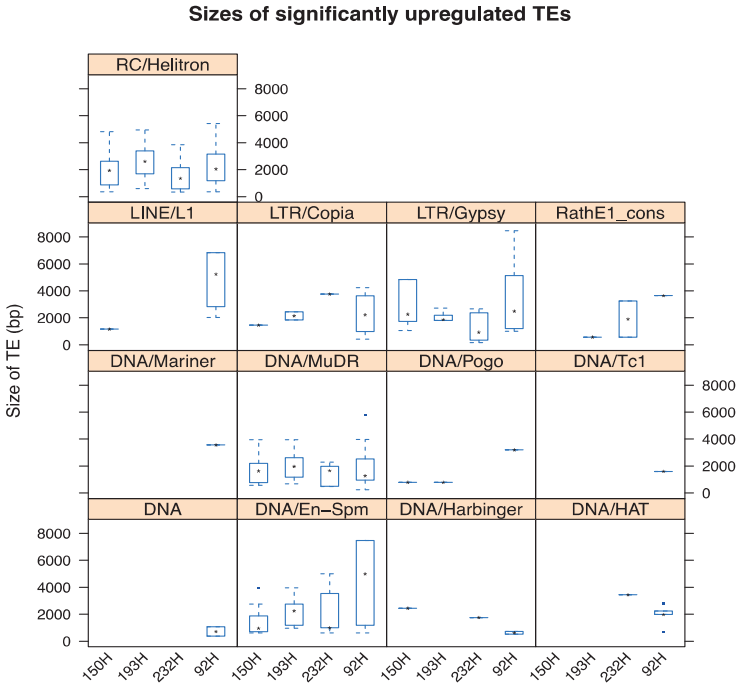

Size of TE (bp)
15000
10000
5000
DNA
DNA/En-Spm
DNA/Harbinger
DNA/HAT
DNA/Mariner
DNA/MuDR
DNA/Pogo
DNA/Tc1
LINE?
LINE/L1
LTR/Copia
LTR/Gypsy
RathE1_cons
RathE2_cons
RathE3_cons
RC/Helitron
SINE
Unassigned

B

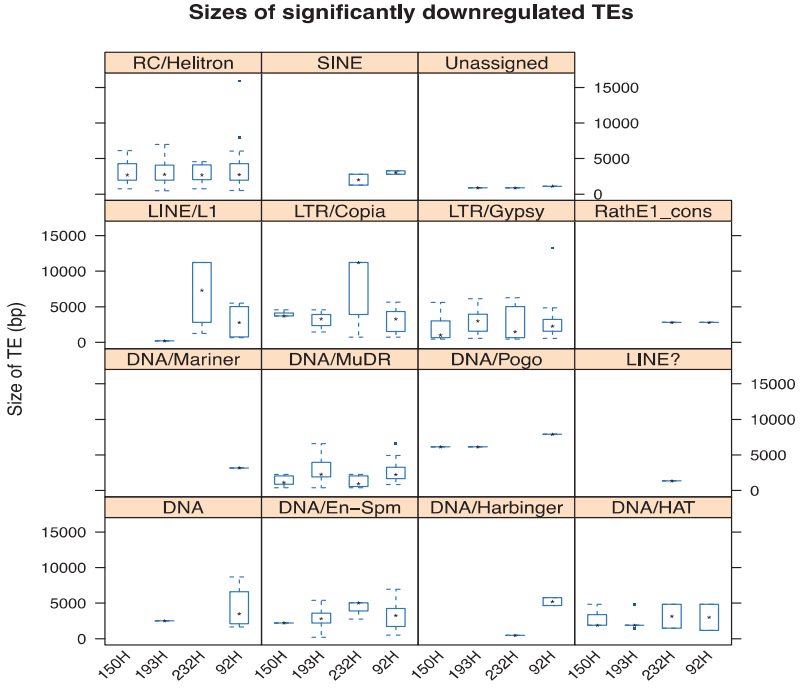

Size of TE (bp)
8000
6000
4000
2000
DNA
DNA/En-Spm
DNA/Harbinger
DNA/HAT
DNA/Mariner
DNA/MuDR
DNA/Pogo
DNA/Tc1
LINE?
LINE/L1
LTR/Copia
LTR/Gypsy
RathE1_cons
RathE2_cons
RathE3_cons
RC/Helitron
SINE
Unassigned

Supplemental Figure 5: Distribution of TE-sizes in the indicated transposon families. (A) all data points and (B) a zoomed-in scale. The grey line indicates 4kb. The dot represents the median.

A


Sizes of significantly upregulated TEs
RC/Helitron
LINE/L1
LTR/Copia
LTR/Gypsy
RathE1_cons
DNA/Mariner
DNA/MuDR
DNA/Pogo
DNA/Tc1
DNA
DNA/En−Spm
DNA/Harbinger
DNA/HAT
Size of TE (bp)
0
2000
4000
6000
8000
150H
193H
232H
92H

B


Sizes of significantly downregulated TEs
RC/Helitron
SINE
Unassigned
LINE/L1
LTR/Copia
LTR/Gypsy
RathE1_cons
DNA/Mariner
DNA/MuDR
DNA/Pogo
LINE?
DNA
DNA/En−Spm
DNA/Harbinger
DNA/HAT
Size of TE (bp)
0
5000
10000
15000
150H
193H
232H
92H

Supplemental Figure 6: Sizes of transposons (in bp) in the indicated families that were (A) upregulated >2 fold from MPV and (B) downregulated >2 fold from MPV in the epiHybrids. The grey line marks 4kb.

Transposon family	Average size all TEs	Average size upregulated TEs in epiHybrids	Average size downregulated TEs in epiHybrids
DNA	317	721	3981
DNA/En-Spm	497	2774	3057
DNA/Harbinger	531	1367	3647
DNA/HAT	383	2366	2690
DNA/Mariner	225	3565	3188
DNA/MuDR	502	1786	2118
DNA/Pogo	368	1594	6721
DNA/Tc1	215	1603	NA
Line?	281	NA	1358
Line/L1	642	4103	4196
LTR/Copia	847	2344	4968
LTR/Gypsy	739	2730	2721
RathE1_cons	129	2012	2769
RathE2_cons	173	NA	NA
RathE3_cons	140	NA	NA
RC-Helitron	539	2185	3219
SINE	246	NA	2537
Unassigned	2231	NA	968

Supplemental Table 2: Average sizes (in bp) of the indicated TE families and TE families upregulated in the epiHybrids.

Supplemental Figure 7: Genome-browser view on DNA methylation (5mC) and RNA expression (RNA) at the AT3G25490 candidate gene from the FT QTL^{epi} region 1 in all lines. This gene is from the selection based on DNA methylation data (SI table 3). DNA methylation and RNA expression is indicated in green and blue, respectively. Scale: 5mC (0 – 0.1); RNA (0 – 50). The red arrow indicates direction of transcription at the candidate gene.

A

	Marker ID			epiRIL			
Interval	ID	Start	Stop	92	150	193	232
FT_1	MM399	8937118	8938596	W	D	W	D
	MM400	9226930	9231459	W	D	W	D
	MM402	9435169	9438146	W	D	D	D
	MM405	9692187	9698029	W	D	D	D
	MM414	11116135	11120705	W	W	D	D
FT_2	MM546	22232528	22235548	W	D	D	D
	MM547	23204533	23207428	W	D	D	D
HT_1	MM695	10992786	10997950	W	W	D	W
	MM698	11363438	11369209	W	W	D	W
	MM699	11820298	11824662	W	W	D	W

B

First choice candidates (same direction in terms of difference between parents and mid-parent divergence.)
Second choice candidates (most cases are a mix of directions in terms of difference between parents or mid-parent divergence)
-/(L or H) Almost significant and/or same direction of mid-parent divergence should match the same haplotype

The haplotype combination (e.g. 92: W, 150: D, 193: W, 232: D) that has a match with the observed mid-parent methylation divergence among the epiHybrids: That means all D's or all W's in this column should match with all L's or all H's in column "3 SD difference from zero".
"W" stands for Col-wt haplotype. "D" stands for *ddm1*-haplotype.
This is relevant because interval FT_1 has different combinations of haplotypes (at the 5 markers within the QTL region) in the epiHybrids (see SI table_3b).

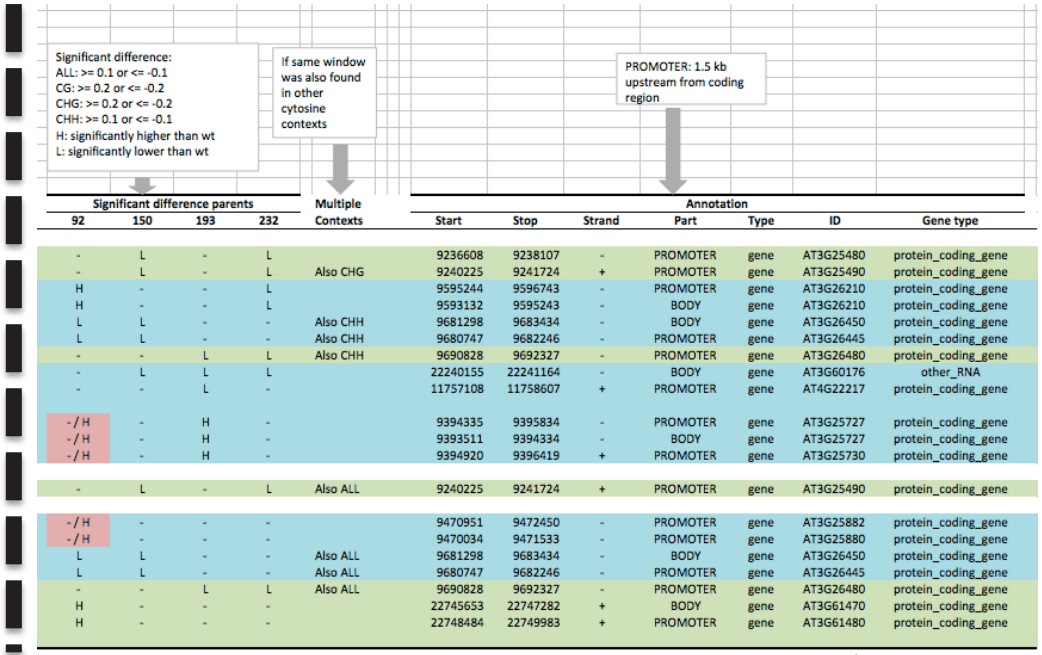
epiHybrid DNA methylation level - mid-parentDNA methylation level

epiHybrid methylation level - mid-parent methylation level expressed in terms of standard deviations from zero (no difference)

Indication if value is lower (L) or higher (H) than 3 standard deviations from zero.

Methylation level epiRIL - methylation level Col-wt

QTL Interval	Cytosine Seq.Context	Match haplotype combination				Window			Methylation epiHybrid - mid-parent value				SD from zero				3 SD difference from zero				Methylation epiRIL - wt			
		92	150	193	232	Chr	Start	Stop	92	150	193	232	92	150	193	232	92	150	193	232	92	150	193	232
FT_1	ALL	W	D	W	D	3	9237501	9238500	0,0144	0,0664	0,017	0,0564	0,86	3,92	1,05	3,35	-	H	-	H	0,0479	-0,2785	0,0415	-0,3163
FT_1	ALL	W	D	W	D	3	9240001	9241000	0,0369	0,0754	0,0288	0,0532	2,2	4,46	1,78	3,16	-	H	-	H	0,033	-0,2674	0,0337	-0,2678
FT_1	ALL	W	D	D	D	3	9595001	9596000	-0,0632	0,0277	0,0693	0,0105	-3,76	1,63	4,29	0,62	L	-	H	-	0,1234	0,0589	0,0679	-0,1225
FT_1	ALL	W	D	D	D	3	9595001	9596000	-0,0632	0,0277	0,0693	0,0105	-3,76	1,63	4,29	0,62	L	-	H	-	0,1234	0,0589	0,0679	-0,1225
FT_1	ALL	W	W	D	D	3	9682001	9683000	0,0811	0,0962	0,0349	-0,1341	4,83	5,68	2,16	-7,97	H	H	-	L	-0,1266	-0,151	-0,0262	-0,0672
FT_1	ALL	W	W	D	D	3	9682001	9683000	0,0811	0,0962	0,0349	-0,1341	4,83	5,68	2,16	-7,97	H	H	-	L	-0,1266	-0,151	-0,0262	-0,0672
FT_1	ALL	W	W	D	D	3	9691501	9692500	-0,0037	-0,0203	-0,0752	-0,0689	-0,22	-1,2	-4,66	-4,09	-	-	L	L	-0,0916	-0,0942	-0,2226	-0,2202
FT_2	ALL	W	D	D	D	3	22239501	22240500	0,0067	0,0593	0,0507	0,0523	0,4	3,5	3,14	3,1	-	H	H	H	0,0198	-0,332	-0,3242	-0,332
HT_1	ALL	W	W	D	W	4	11756501	11757500	0,0227	-0,0516	0,0974	0,0431	1,35	-3,05	6,03	2,56	-	L	H	-	-0,0414	0,0344	-0,1357	-0,0899
FT_1	CG	W	D	W	D	3	9394001	9395000	-0,0737	0,0688	-0,1084	0,0063	-2,02	1,85	-3,09	0,17	-/L	-	L	-	0,1989	0,0654	0,2266	9,00E-04
FT_2	CG	W	D	W	D	3	9394001	9395000	-0,0737	0,0688	-0,1084	0,0063	-2,02	1,85	-3,09	0,17	-/L	-	L	-	0,1989	0,0654	0,2266	9,00E-04
FT_3	CG	W	D	W	D	3	9394001	9395000	-0,0737	0,0688	-0,1084	0,0063	-2,02	1,85	-3,09	0,17	-/L	-	L	-	0,1989	0,0654	0,2266	9,00E-04
FT_1	CHG	W	D	W	D	3	9240001	9241000	0,0454	0,1179	0,0347	0,1035	1,56	4,14	1,31	3,52	-	H	-	H	0,096	-0,4645	0,0622	-0,4642
FT_1	CHH	W	D	D	D	3	9471501	9472500	-0,0593	0,0475	0,0339	-0,0142	-4,14	3,29	2,35	-0,98	L	H	-	-	0,0947	-0,0235	-0,0124	-0,0267
FT_1	CHH	W	D	D	D	3	9471501	9472500	-0,0593	0,0475	0,0339	-0,0142	-4,14	3,29	2,35	-0,98	L	H	-	-	0,0947	-0,0235	-0,0124	-0,0267
FT_1	CHH	W	W	D	D	3	9682001	9683000	0,0651	0,0901	NA	-0,1354	4,55	6,25	NA	-9,37	H	H	-	L	-0,1286	-0,1393	NA	-0,0744
FT_1	CHH	W	W	D	D	3	9682001	9683000	0,0651	0,0901	NA	-0,1354	4,55	6,25	NA	-9,37	H	H	-	L	-0,1286	-0,1393	NA	-0,0744
FT_1	CHH	W	W	D	D	3	9692001	9693000	0,0056	-0,0076	-0,0438	-0,0667	0,39	-0,52	-3,04	-4,61	-	-	L	L	-0,0746	-0,0586	-0,2907	-0,291
FT_2	CHH	W	D	D	D	3	22747001	22748000	0,043	-0,0037	0,0016	0,0109	3	-0,25	0,11	0,76	H	-	-	-	0,1089	0,0046	0,0213	-0,0023
FT_2	CHH	W	D	D	D	3	22747501	22748500	0,0536	0,0057	0,0257	0,0253	3,74	0,4	1,78	1,75	H	-	-	-	0,1396	-0,0019	0,0264	-0,0106



Significant difference:
ALL: >= 0.1 or <= -0.1
CG: >= 0.2 or <= -0.2
CHG: >= 0.2 or <= -0.2
CHH: >= 0.1 or <= -0.1
H: significantly higher than wt
L: significantly lower than wt

If same window was also found in other cytosine contexts

PROMOTER: 1.5 kb upstream from coding region

Significant difference parents				Multiple	Annotation						
92	150	193	232	Contexts	Start	Stop	Strand	Part	Type	ID	Gene type
-	L	-	L		9236608	9238107	-	PROMOTER	gene	AT3G25480	protein_coding_gene
-	L	-	L	Also CHG	9240225	9241724	+	PROMOTER	gene	AT3G25490	protein_coding_gene
H	-	-	L		9595244	9596743	-	PROMOTER	gene	AT3G26210	protein_coding_gene
H	-	-	L		9593132	9595243	-	BODY	gene	AT3G26210	protein_coding_gene
L	L	-	-	Also CHH	9681298	9683434	-	BODY	gene	AT3G26450	protein_coding_gene
L	L	-	-	Also CHH	9680747	9682246	-	PROMOTER	gene	AT3G26445	protein_coding_gene
-	-	L	L	Also CHH	9690828	9692327	-	PROMOTER	gene	AT3G26480	protein_coding_gene
-	L	L	L		22240155	22241164	-	BODY	gene	AT3G60176	other_RNA
-	-	L	-		11757108	11758607	+	PROMOTER	gene	AT4G22217	protein_coding_gene
- / H	-	H	-		9394335	9395834	-	PROMOTER	gene	AT3G25727	protein_coding_gene
- / H	-	H	-		9393511	9394334	-	BODY	gene	AT3G25727	protein_coding_gene
- / H	-	H	-		9394920	9396419	+	PROMOTER	gene	AT3G25730	protein_coding_gene
-	L	-	L	Also ALL	9240225	9241724	+	PROMOTER	gene	AT3G25490	protein_coding_gene
- / H	-	-	-		9470951	9472450	-	PROMOTER	gene	AT3G25882	protein_coding_gene
- / H	-	-	-		9470034	9471533	-	PROMOTER	gene	AT3G25880	protein_coding_gene
L	L	-	-	Also ALL	9681298	9683434	-	BODY	gene	AT3G26450	protein_coding_gene
L	L	-	-	Also ALL	9680747	9682246	-	PROMOTER	gene	AT3G26445	protein_coding_gene
-	-	L	L	Also ALL	9690828	9692327	-	PROMOTER	gene	AT3G26480	protein_coding_gene
H	-	-	-		22745653	22747282	+	BODY	gene	AT3G61470	protein_coding_gene
H	-	-	-		22748484	22749983	+	PROMOTER	gene	AT3G61480	protein_coding_gene

Supplemental Table 3: A) shows the haplotypes at the respective QTL^{epi} regions and B) Candidate selection within the QTL^{epi} regions based on DNA methylation divergence. FT_ 1 and 2 and HT_1 refer to two QTL^{epi} regions for flowering time on Chr. 3 and the single QTL^{epi} region for plant height on Chr. 4 (Chapter 3).

	Start	End	Gene	Chr.	QTL start	QTL end	epiQTL Interval	Description
92H								
signif. upregulated								
	9184688	9184792	AT3G25221	3	8937118	9698020	FT_1	unknown protein;
	22291597	22291737	AT3G60318	3	22232528	23207428	FT_2	This gene encodes a small protein and has either evidence of transcription or purifying selection.
	22925813	22926379	AT3G61900	3	22232528	23207428	FT_2	SAUR-like auxin-responsive protein family
	22932496	22933362	AT3G61920	3	22232528	23207428	FT_2	unknown protein;
	22971358	22972268	AT3G62020	3	22232528	23207428	FT_2	germin-like protein (GLP10)
	23087017	23089206	AT3G62390	3	22232528	23207428	FT_2	Encodes a member of the TBL (TRICHOME BIREFRINGENCE-LIKE) gene family
	23135984	23137173	AT3G62550	3	22232528	23207428	FT_2	Adenine nucleotide alpha hydrolases-like superfamily protein; INVOLVED IN: response to stress;
	11424710	11425767	AT4G21445	4	11363438	11824662	HT_1	unknown protein;
signif. downregulated								
	9110103	9112748	AT3G25010	3	8937118	9698020	FT_1	Receptor like protein 41 (RLP41); FUNCTIONS IN: kinase activity; INVOLVED IN: signal transduction, defense response;
	9116868	9119540	AT3G25020	3	8937118	9698020	FT_1	Receptor like protein 42 (RLP42); INVOLVED IN: signal transduction, defense response;
	9308942	9313353	AT3G25610	3	8937118	9698020	FT_1	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein;
	9409290	9410567	AT3G25780	3	8937118	9698020	FT_1	Encodes allene oxide cyclase, one of the enzymes involved in jasmonic acid biosynthesis. mRNA expression is upregulated in senescing leaves.
	9676564	9679900	AT3G26440	3	8937118	9698020	FT_1	Protein of unknown function (DUF707);
	9687273	9690827	AT3G26480	3	8937118	9698020	FT_1	Transducin family protein / WD-40 repeat family protein;
	22334424	22337617	AT3G60420	3	22232528	23207428	FT_2	Phosphoglycerate mutase family protein;
	22374963	22375865	AT3G60540	3	22232528	23207428	FT_2	Preprotein translocase Sec, Sec61-beta subunit protein;
	22650159	22651488	AT3G61190	3	22232528	23207428	FT_2	Encodes a protein with a C2 domain that binds to BON1 in yeast two hybrid analyses. Involved in maintaining cell homeostasis.
	23008755	23013909	AT3G62150	3	22232528	23207428	FT_2	P-glycoprotein 21 (PGP21); FUNCTIONS IN: ATPase activity, coupled to transmembrane movement of substances; INVOLVED IN: transport, transmembrane transport;
	11587049	11588272	AT4G21840	4	11363438	11824662	HT_1	methionine sulfoxide reductase B8 (MSRB8); FUNCTIONS IN: peptide-methionine-(S)-S-oxide reductase activity; INVOLVED IN: oxidation reduction;
	11591116	11592238	AT4G21850	4	11363438	11824662	HT_1	methionine sulfoxide reductase B9 (MSRB9); FUNCTIONS IN: peptide-methionine-(S)-S-oxide reductase activity; INVOLVED IN: oxidation reduction;
150H								
signif. upregulated								
	22398228	22399904	AT3G60590	3	22232528	23207428	FT_2	unknown protein;
signif. downregulated								
	9099183	9101837	AT3G24900	3	8937118	9698020	FT_1	Receptor like protein 39 (RLP39); INVOLVED IN: signal transduction, defense response;
	9110103	9112748	AT3G25010	3	8937118	9698020	FT_1	Receptor like protein 41 (RLP41); FUNCTIONS IN: kinase activity; INVOLVED IN: signal transduction, defense response;
	9116868	9119540	AT3G25020	3	8937118	9698020	FT_1	Receptor like protein 42 (RLP42); INVOLVED IN: signal transduction, defense response;
	9676564	9679900	AT3G26440	3	8937118	9698020	FT_1	Protein of unknown function (DUF707); CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF707
	11587049	11588272	AT4G21840	4	11363438	11824662	HT_1	methionine sulfoxide reductase B8 (MSRB8); FUNCTIONS IN: peptide-methionine-(S)-S-oxide reductase activity; INVOLVED IN: oxidation reduction;
	11644218	11644605	AT4G21950	4	11363438	11824662	HT_1	unknown protein;

193H								
signif. upregulated								
	22398228	22399904	AT3G60590	3	22232528	23207428	FT_2	unknown protein;
	22971358	22972268	AT3G62020	3	22232528	23207428	FT_2	germin-like protein (GLP10)
	23135984	23137173	AT3G62550	3	22232528	23207428	FT_2	Adenine nucleotide alpha hydrolases-like superfamily protein; INVOLVED IN: response to stress;
signif. downregulated								
	11535525	11542268	AT4G21710	4	11363438	11824662	HT_1	Encodes the unique second-largest subunit of DNA-dependent RNA polymerase II;
	11644218	11644605	AT4G21950	4	11363438	11824662	HT_1	unknown protein;
232H								
signif. upregulated								
	22398228	22399904	AT3G60590	3	22232528	23207428	FT_2	unknown protein;
	22929006	22930442	AT3G61910	3	22232528	23207428	FT_2	NAC transcription factor NST2. NST1 and NST2 are redundant in regulating secondary wall thickening in anther walls.
	23135984	23137173	AT3G62550	3	22232528	23207428	FT_2	Adenine nucleotide alpha hydrolases-like superfamily protein;
	11601905	11603276	AT4G21865	4	11363438	11824662	HT_1	unknown protein;
	11653380	11654136	AT4G21970	4	11363438	11824662	HT_1	Protein of unknown function, DUF584;
	11657008	11659167	AT4G21990	4	11363438	11824662	HT_1	Encodes a protein disulfide isomerase-like (PDIL) protein, a member of a multigene family within the thioredoxin (TRX) superfamily.
signif. downregulated								
	9110103	9112748	AT3G25010	3	8937118	9698020	FT_1	Receptor like protein 41 (RLP41); FUNCTIONS IN: kinase activity; INVOLVED IN: signal transduction, defense response;
	9113168	9113281	AT3G25011	3	8937118	9698020	FT_1	unknown protein;
	9114546	9114707	AT3G25014	3	8937118	9698020	FT_1	unknown protein;
	9116868	9119540	AT3G25020	3	8937118	9698020	FT_1	Receptor like protein 42 (RLP42); INVOLVED IN: signal transduction, defense response;
	22541319	22552529	AT3G60961	3	22232528	23207428	FT_2	P-loop containing nucleoside triphosphate hydrolases superfamily protein; FUNCTIONS IN: nucleobase, nucleoside, nucleotide kinase activity, ATP binding;
	23146828	23147951	AT3G62580	3	22232528	23207428	FT_2	Late embryogenesis abundant protein (LEA) family protein;
	11587049	11588272	AT4G21840	4	11363438	11824662	HT_1	methionine sulfoxide reductase B8 (MSRB8); FUNCTIONS IN: peptide-methionine-(S)-S-oxide reductase activity; INVOLVED IN: oxidation reduction;
	11591116	11592238	AT4G21850	4	11363438	11824662	HT_1	methionine sulfoxide reductase B9 (MSRB9); FUNCTIONS IN: peptide-methionine-(S)-S-oxide reductase activity; INVOLVED IN: oxidation reduction;
	11644218	11644605	AT4G21950	4	11363438	11824662	HT_1	unknown protein;

Supplemental Table 4: Genes with significant (> 2 fold up/downregulation) from MPV in the QTL^{epi} regions. Chr. stands for chromosome. FT_ 1 and 2 and HT_1 refer to two QTL^{epi} regions for flowering time on Chr. 3 and the single QTL^{epi} region for plant height on Chr. 4 (Chapter 3). Genes that are shared among two or more hybrids are highlighted with the same background color.