Phenotypic variation in plants

Roles for epigenetics

Lauss, K.

Publication date
2017
Document Version
Other version
License
Other

Chapter 6

General Discussion

DNA sequence variation is a major driver of phenotypic variation [196]. However, with genomes and their epigenetic modifications becoming more thoroughly characterized, the contribution of epigenetics to phenotypic diversity is being better understood and is gaining in appreciation. Epigenetics encompasses modifications of the DNA sequence and chromatin that can affect gene activity and genome function [6,7] and represents an additional layer of information beside the genetic code.

In this thesis we aim to contribute to an understanding of the relationship between the phenotype and epigenome, focusing pre-dominantly on DNA methylation, in two main projects. In the first one, we investigated the contribution of epigenetic divergence in parental lines to heterosis in various traits in Arabidopsis hybrids. We combined a phenotypic study in hybrids and their parental lines with a recently developed method of mapping epigenetic quantitative trait loci (**Chapters 2** and **3**) and with genome-wide analysis of DNA methylation and RNA expression in a subset of the lines (**Chapter 4**).

In the second project, we investigated the mechanism underlying paramutation, an epigenetic phenomenon where *in trans* communication between DNA regions with contrasting epigenetic profiles results in one region heritably acquiring the profile of the other (**Chapter 5**). Paramutation can result in non-Mendelian inheritance patterns of traits. We studied paramutation at two epialleles of the maize *b1* locus, a repressed epiallele (*B'*) and an active epiallele (*B-I*), which results in plants displaying low or high anthocyanin pigmentation levels, respectively. Upon crossing (hybridizing) lines carrying *B'* and *B-I*, *B-I* heritably acquires the epigenetic profile of *B'* resulting in the light pigmentation phenotype in the progeny, in 100% of the cases. We studied DNA methylation and small regulatory RNAs in early steps of the paramutation process in an attempt

to gain insights into relevant features of the alleles involved and the early sequence of events.

In this chapter we discuss the topics and findings of this thesis.

Studying the contribution of epigenetic modifications to phenotypic variation:

epiRILs & QTL^{epi} mapping

Studying the epigenetic contribution to phenotypic diversity in natural populations is challenging as epigenetic variation is usually accompanied by DNA sequence variation [52,53]. Disentangling whether factors responsible for a given phenotype in such lines are epigenetic or genetic is difficult. Two populations of epigenetically recombinant inbred lines (epiRILs) of Arabidopsis have been instrumental in overcoming this challenge. The epiRIL populations were derived from crosses between a mutant defective in DNA methylation, *ddm1* (*decreased In DNA methylation 1*) or *met1* (*methyltransferase 1*), and its respective wildtype plant (Col-0 wildtype) [52,53]. Rounds of back-crossing and selfing (up to the F8) resulted in lines that have mosaic epigenomes that are virtually homozygous for the methylation states originating from either the wild-type parent or the mutant parent. Since epiRILs have only marginal genetic differences (caused mostly by re-activated transposons that lost their silencing marks) [52,53,56,57,126,138], but substantial differences in their DNA methylation profiles, they are very useful for assessing the contribution of DNA methylation variation to plant phenotypes. Phenotypic analyses of the epiRIL populations demonstrated widespread variation for traits like plant height, biomass, fruit (silique) size, number of fruits, time to flowering, germination rates, and response to bacterial infection [52,53,55].

One big question that remains open is how much of the phenotypic variation in plant populations is a consequence of epigenetic rather than

genetic contributions. The contribution of genetic factors is generally identified by classical quantitative trait loci (QTL) mapping approaches [197]. Classic QTL mapping may completely miss epigenetic quantitative trait loci ($QTLs^{epi}$) or wrongly assign a QTL^{epi} to a genetic polymorphism instead of an epigenetic one. Therefore, epiRILs have been used in one of the first successful attempts to map $QTLs^{epi}$ [56,57]. In Chapter 2, we describe the steps of $QTLs^{epi}$ mapping, starting from the creation of epiRILs up to the use of stably inherited DNA methylation markers to create linkage maps, and the validation of identified candidate regions and subsequently genes. Challenges of QTL^{epi} mapping include the identification of stably inherited epigenetic markers [106]. The stability of epigenetic markers is illustrated by the rate of epimutations. Epimutations in general describe heritable changes in DNA methylation or in the even more dynamic histone modifications at a given position in the DNA sequence. With regard to DNA methylation, the term epimutation refers to spontaneous gains or losses of DNA methylation at single cytosine residues or DNA regions [198]. Epimutations happen predominantly in CG sequence context, with more CG-methylation divergence in gene-rich regions than in pericentromeric Transposable Element (TE)-rich regions [199–201]. For QTL^{epi} mapping, the stability of methylation markers needs to be assessed before creating a linkage map (i.e. by comparing DNA methylation markers in the parental lines with those in the mapping population) and the stability of the markers needs to be confirmed for the mapping population of interest. The most recently determined epimutation rates (per CG site, per haploid methylome, per generation) in Arabidopsis are on average about five orders of magnitude higher than the genetic mutation rate (7×10^{-9}) [202], whereby the rate of methylation gain (3×10^{-4}) is slightly lower than the rate of methylation loss (6×10^{-4}) [203]. DNA regions of contiguous methylation (with a length between 50

bp and 650 bp) have epimutation rates comparable to the genetic mutation rate and are hence more stable than single CG sites [199,202]. For QTL^{epi} mapping in Arabidopsis, differentially methylated regions (DMRs) between lines have been used as markers, rather than single cytosines [56].

Another challenge of QTL^{epi} mapping lies in confirming that differences in epigenetic profiles at certain regions are causal for a particular phenotype. This is complicated by the fact that variation in DNA methylation can occur in ways that are dependent on, or independent of DNA sequence variation [142,163]. If DNA methylation in a putative QTL^{epi} region is linked to a single-nucleotide polymorphism (SNP) or a structural variant (i.e. insertions, deletions, translocations), then both the region's genetic and the region's epigenetic profile can be relevant for the quantitative trait locus. In standard QTL mapping using genetic markers, one strategy to confirm the presence of a QTL region is creating and phenotyping a nearly isogenic line (NIL) harboring the QTL interval of genome "A" in the background of genome "B". In principle, the same approach could be taken for a QTL^{epi}, which can be traced by epigenotyping of markers that are present in the region of interest. To further confirm the causality of particular methylation profiles at selected loci, transgenes can be introduced that induce the production of small RNAs, which in turn induce DNA methylation at their target site [114,115]. Also, rapidly developing techniques for site-specific epigenomic editing can be applied. Such methods are based on catalytic domains of chromatin- or DNA-modifying enzymes, fused to a specific DNA recognition domain, such as CRISPR-Cas, Zinc finger proteins or transcription-activator-like effectors (TALEs) [116,204,205]. Such techniques enable targeting of the desired chromatin modification to any locus of interest [116,204,205].

Despite aforementioned challenges, QTL^{epi} mapping is a big step forward in understanding the contribution of epigenetic variation to phenotypic variation.

Epigenomic differences between parental lines can drive heterosis

Heterosis is an agriculturally important phenomenon that refers to hybrid progeny outperforming its parents in a trait of interest [63,67,71]. Various genetic models have been proposed to explain heterosis [69,70,73–75], but rather recent findings indicated that, besides differences in genetic background, also differences in the parental epigenome contribute to heterosis [45,47,48,72,77,103,132]. Importantly, these observations have been confirmed in an Arabidopsis epiHybrid displaying heterosis for biomass that was generated by crossing a *met1*-derived epiRIL with wildtype Col-0 (Col-wt) [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 extended the view on the role of epigenetic mechanisms in heterosis by creating epiHybrids using a selection of lines from the *ddm1*-derived epiRILs. We created F1 epiHybrids by crossing Col-wt as a maternal parent to 19 different epiRILs and we measured various traits including leaf area (LA), flowering time (FT) and final plant height (HT). The reason for performing the crosses with Col-wt as maternal parent was to keep the cytoplasmic components similar in all crosses. We observed several strong, positive and negative heterotic phenotypes among the epiHybrids, suggesting that epigenetic divergence is sufficient to cause heterosis. Heterosis for FT and HT could be associated with two and one putative QTL^{epi} regions in the parental genomes, respectively. For LA, there were indications for the same two QTL^{epi} regions as for FT, however, for LA they remained below the significance threshold. The

DMRs in these regions contained 35 (in the two FT regions) and 14 (in the one HT region) genes whose differential methylation profiles may underlie the heterotic phenotypes observed. Our study contrasts with the previously mentioned report from Dapp *et al.* [35] in the sense that: a) we observed a range of phenotypic effects while using a similar maternal cytoplasm (Col-wt) for all crosses, implying we monitored effects linked to specific epiRIL methylomes rather than to cytoplasmic factors; b) we examined multiple (nineteen) rather than one single epiHybrid, allowing us to perform a proof-of-principle of the generality of the QTL^{epi} mapping approach. Thereby, we could link some of the heterotic phenotypic effects to methylation differences at particular regions in the parental genomes. In sum, our results, more massively than those of Dapp *et al.* [58], indicate that epigenetic differences can be sufficient to cause heterosis. To get an idea of the molecular mechanisms underlying our heterotic phenotypes we proceeded with analyzing the DNA methylomes and RNA transcriptomes of epiHybrids and their parental lines.

Preliminary analysis of genes at the QTL^{epi} regions

To get insights into the potential molecular mechanisms underlying our heterosis $QTLs^{epi}$ we performed a rough and very preliminary selection based on non-additive DNA methylation and gene expression changes on the genes within the QTL^{epi} intervals (Chapter 4). We found that the candidate genes identified by selecting for significant non-additive methylation changes in promoter or gene body did not overlap with candidate genes that displayed non-additive expression changes. We concluded that the approach to look first for interesting (non-additive) gene expression patterns and only then investigate DNA methylation at the gene promoter to be more fruitful than *vice versa* (Chapter 4).

From a molecular perspective, very specific expression patterns can play a

role in heterosis and this can also involve other than just non-additive changes. Gene expression levels around mid-parent values can also be relevant, for instance, when dosage-dependent regulation is concerned. A dosage effect concerning yield heterosis in tomato has been described: the SINGLE FLOWER TRUSS (SFT) gene is in a heterozygous (overdominant) state expressed at between-parental levels and affects a pathway that is sensitive to dosage causing the heterotic phenotype [158,159]. Also expression of SA-related genes to in-between parent levels has been proposed to mediate seedling growth heterosis in Arabidopsis [153]. We identified an oxidative stress response gene in HT QTL^{epi} were such dosage effect might play a role for the heterotic height phenotype observed: the gene showed significant negative mid-parent divergence in epiHybrids 92H, 150H and 232H - but displayed an expression level closer to mid-parent value in 193H. And 193H was the only hybrid that displayed best-parent heterosis (BPH) for the plant height trait.

In sum, we believe that additive expression of candidate genes (or combinations of candidate genes) from the QTL^{epi} regions may just as likely underlie heterotic phenotypes as non-additive changes.

Elusive links between gene expression and DNA methylation

DNA methylation at genes can be present both in promoter regions and in the gene body [140]. Promoter methylation commonly results in reduced gene transcription, while gene-body methylation has been associated with moderate-to-high gene expression [22,23].

Our preliminary attempts to study candidate genes within the QTLs^{epi} identified in Chapter 3, showed that associating non-additive expression changes with according DNA methylation changes at promoter regions was challenging. We speculated that more distant DMRs, i.e. at regulatory elements elsewhere within the QTLs^{epi} or histone modifications associated

with DMRs could also cause effects on gene expression and would not be detected by studying only promoter methylation. A *ddm1* mutation (present in the epiRIL founder parent) does not only reduce DNA methylation but results also in increases of activating histone methylation marks (H3K4me1/2/3) and decreases of repressive histone modifications (H3K9me2) [30,206]. This may also play a role in the altered transcript levels of *ddm1* mutants and their descendants. An involvement of histone modifications in regulating genes involved in heterosis has recently been proposed in a study describing the effects of DDM1 on heterosis in Arabidopsis: SA-related genes were stronger expressed in a hybrid in *ddm1* mutant background. This hybrid displayed reduced heterosis than did the hybrid in a non-mutant background. However, the genes affected displayed no or only low DNA methylation in both the Col-wt and *ddm1* mutant. The authors [153] put forward that , histone modification rather than DNA methylations was involved in the effect and SA-related genes.

Other insights into the relationship between DNA methylation and gene expression come from genome-wide studies of transcript level dynamics in a *met1*-derived epiHybrid: Rigal et al. found 1576 genes upregulated in the *met1* mutant parent compared to wildtype, indicating that substantial loss of methylation enhances expression of gene regions [78]. However, also here the link between methylation and expression was subtle at best: Of the 1576 genes, 40% had intermediate expression levels and 40% were repressed in the epiHybrid [78]. Interestingly, the transcription repression of the latter 40% was independent of DNA re-methylation. The authors reported only two loci that were an exception to this rule: the genes SDC and KELP where silenced in their epiHybrid and this correlated with DNA re-methylation of tandem repeats located in their promoter regions [78]. These results suggest that expression of a random gene cannot solely be inferred from its DNA methylation state.

In sum, at this stage we could not determine which exact expression and DNA methylation patterns at the QTLs^{epi} were causal for the heterotic phenotypes we observed. Fine-mapping of the QTL^{epi} regions (as described in Chapter 2) might be necessary to reveal causal relationships between expression and methylome profiles.

Trans-chromosomal (de)methylation events in epiHybrids

Crossing parental lines with a different (epi)genomic backgrounds to create an F1 hybrid results in the combination of two different genomes and epigenomes. Epigenetic profiles or more specifically the two DNA methylomes are shown to interact in hybrids in the form of trans-chromosomal (de)methylation (TC(d)M) events [78,79,146]. TC(d)M events were demonstrated in genetic as well as in a *met1*-derived epigenetic hybrids. The corresponding studies describe *in trans* communication between differentially or similarly methylated regions resulting in identical sequence regions acquiring a similar methylation profile [78,79,146]. This can lead to a gain but also a loss of DNA methylation. Small RNAs have been suggested as mediators of such methylome interactions [78,79,146] in a pathway termed RNA-dependent DNA methylation (RdDM). 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].

In Chapter 4, we found indications for TC(d)M events in all four *ddm1*-derived epiHybrids studied, with loss of DNA methylation occurring in 0.5 to 1.3% of all genomic windows (sliding window using a size of 1 kb and steps of 500 bp) and gain of methylation in 1.2 to 2% of all windows. Most TCdM events occurred at regions that were differentially methylated in the parents, but the majority of TCM events appeared to occur at regions that did not show significant methylation differences between the parents. Both observations are consistent with previous reports on hybrids

between Arabidopsis accessions [79,146]. This indicates that a large fraction of TCdM interactions are allelic while TCM events may also occur frequently between non-allelic regions. We found only few TCdM and TCM events that occurred in all four epiHybrids. The largest fractions of TC(d)M events were unique to the individual epiHybrids, which may indicate that the specific DNA methylomes of the epiRIL parents strongly influence methylome interactions in the epiHybrids.

With respect to the role of methylome interactions in heterosis, a few observations have been reported: a non-functional RdDM pathway in Arabidopsis does not disrupt heterosis in leaf rosette radius in 20 day old seedlings [146]. Similarly, despite a mutation in the RdDM pathway and corresponding reduction in small RNAs in maize hybrids (between B73 and Mo17), heterosis in cob weight of maize hybrids was not disrupted either [45]. In an epiHybrid generated from a *met1* mutant and Col-wt, thousands of regions undergoing TC(d)M have been identified, but the epigenetic hybrid did not display biomass heterosis [78]. This suggests that TC(d)M events in general are not essential for heterosis, at least not in the instances studied.

A TC(d)M event, which causes a heritable change of the DNA methylation profile at a given locus, falls under the definition of paramutation. In Arabidopsis hybrids, a few methylation patterns acquired through TCM have been followed to the F2 generation and most of them where indeed stably inherited. Thus, such events show (at least partially) paramutation-like behavior [121,146]. One of those stably inherited TCM events was not detected F1 seedlings but was only identified later in development (F1 floral buds) [121]. This suggests that the choice of plant tissue used for analysis can affect which TC(d)M events are detected. Also, TC(d)M events related to certain phenotypic traits may only be detected later in development.

In sum, preventing TC(d)M may not disrupt heterosis of certain traits. However, TC(d)M events at specific genome locations may still play a role in heterosis in particular traits or in other hybrids than the ones that have been tested to date. How this genome-wide occurrence of TC(d)M affects gene expression and relates to phenotypic variation needs to be addressed by future research efforts.

Genetic models and network-based models that have been proposed to explain heterosis

How exactly our putative QTL^{epi} regions contribute to the heterotic phenotypes observed remains to be investigated. Many genetic explanations have been suggested to explain heterosis and in principle those may hold for stable epialleles as well.

The classical models are the dominance and overdominance hypotheses. The dominance hypothesis postulates that heterosis in a hybrid arises due to complementation of deleterious mutations that are different between the (inbred) parents [69,73,74]. These conclusions have been drawn mainly from studies in maize crops where parental lines were highly inbred, which caused homozygosity at deleterious alleles. This type of heterosis can in principle be fixed in inbred lines by substituting deleterious alleles with good ones.

The overdominance model states that heterozygosity per se at particular alleles can be advantageous (conferring higher fitness) over homozygosity [69,70], resulting in heterosis [69,70]. Creating inbred lines, in which heterosis caused by overdominance is fixated, is difficult as one would need to maintain a heterozygous state at the relevant locus. When more loci are involved, this will become cumbersome if not impossible.

Later, other models have extended the classic explanations of heterosis. One of those is called pseudo-overdominance, and this refers to cases

where heterosis is mediated by heterozygosity of tightly linked loci harboring complementing loci on the opposite homologues. Lastly, another model, the epistasis model, proposes that non-allelic gene interactions contribute to heterosis [75].

There are cases of heterosis reported that have a multi-gene architecture and cases that are linked to single genes. One example for single-gene heterosis was described at a tomato gene, the SINGLE FLOWER TRUSS (SFT), which in a heterozygous (overdominant) state results in yield increases compared to the parental lines [158]. SFT encodes the flowering hormone florigen and the heterosis effect is due to a dosage sensitivity in the florigen pathway [159]. The overdominant effect could even be fixed in inbred lines by introducing artificial miRNAs causing downregulation of SFT. The latter provides an example of how an overdominant effect can be recapitulated in an inbred by interfering with the regulation of gene expression rather than with the genes or their epigenetic status themselves [159]. There is evidence for each of the above hypotheses, suggesting that multiple mechanisms may apply simultaneously or that mechanisms underlying heterosis differ depending on the trait or the species studied.

Beside quantitative genetics approaches (as described in this thesis), there are systems biology approaches that attempted to extend the classical genetic explanations of heterosis with network-based models [207,208]. Approaches in this direction focused mainly on regulatory interactions and multigenic effects, essentially pursuing the idea that favorable epistatic interactions are underlying heterosis [207,208]. For heterosis in Arabidopsis hybrids, a network model has been simulated and even validated using experimental data [207]. The model is based on the idea that heterozygotes, due to a greater diversity in their alleles, possess a higher biochemical versatility (are more adaptable) [207,209]. So, the core

hypothesis of this model is that the superior performance of (heterozygous) hybrids is due to an increase in regulatory interactions compared to in (homozygous) parents, increasing the adaptability to changing environments [207,209]. Experimental data on partial correlations were used to estimate interactions amongst time dependent experimental gene expression and metabolite profiles. Indeed, the analysis demonstrated that network interconnections are denser in hybrids [207,208]. The idea of heterosis also being influenced by increased interactions in gene networks, rather than only being mediated by local effects on particular regions of the DNA, is an interesting one. Similarly, increased epigenetic interactions in a hybrid [46,78,79,146], could play a role in heterosis. Future research might be able to integrate quantitative genetics approaches (QTL mapping) with such network hypotheses, for instance, by studying enriched interactions on multiple OMICS levels at identified QTLs.

At this point it is not really possible to assign genes from within our QTLs^{epi} regions to certain models. To understand heterosis fully it may be also necessary to look beyond these models towards more integrative approaches.

A genetic basis for DNA methylation variation?

Various research efforts have been made to understand whether there is a genetic basis for heritability of DNA methylation variation [113,140,210,211]. In plants, it has been suggested that some of the detected *cis* associations between a particular DNA sequence and DNA methylation are due to single-nucleotide polymorphisms (SNPs) [113,140,210,211]. One possible way in which a SNP may be causing DNA methylation variation is that the SNP could be associated with nearby repeats or transposons (TEs) [113,155,198]. It is known that TEs can

induce gene expression changes, for instance, by inducing spreading of DNA methylation from transposons into flanking genes, by containing regulatory sequences or by small RNAs derived from a TE that regulate host genes [155].

Interestingly, in one of the most recent genome-wide analyses on Arabidopsis accessions only 18-35% of all DMRs could be linked to genetic variation [203]. So, the majority of the DMRs in that study appeared not to be associated with genetic variation. These unassociated DMRs could result from sequence-independent segregation of methylation states [212]. Alternatively, as the rate of epimutations at single cytosine sites is higher than the genetic mutation rate, one may speculate about the possibility that DMRs might have become disassociated from their causative DNA sequence through the course of evolution [198,203]. In Chapter 3 we surmised that the QTLs^{epi} we detected for heterosis in flowering time/leaf area and plant height were not due to shared structural variants (such as insertions, deletions or translocations) in the parental epiRIL lines, meaning that, for example, shared TE insertions in the parental lines are unlikely to be the cause of our detection of QTLs^{epi}. However, as we do not have sequence information (yet) of the QTL^{epi} regions in the epiHybrids, it remains possible that DNA methylation variation in the three QTL^{epi} regions is associated with SNPs. It has been argued that such *cis*-SNP associations could be in a mere linkage disequilibrium with a DMR rather than being causal; this would suggest a genetic basis for variation in DNA methylation were there is no true basis [198]. Also, as described above, not all DNA methylation variation has a genetic basis [203]. The studies of Cortijo et al. and Kooke et al. provide evidence for the existence of pure epialleles regulating trait variation in Arabidopsis [56,57].

Investigating whether DNA methylation variation in the QTLs^{epi} co-occurs with particular SNPs in the epiHybrids is important for drawing more definite conclusions on the role of DNA methylation in heterosis.

Maize as a model: Paramutation at the anthocyanin-pigmentation loci

Besides being an interesting model organism for research, maize is the main cereal grain cultivated throughout the world [213]. It is one of the major examples where heterotic hybrids substantially increased crop yields [71]. Another trait that is interesting for breeding programs is anthocyanin production in maize kernels [214]. Anthocyanins are secondary metabolites that belong to a class of flavonoids that are synthesized by a complex metabolic pathway consisting of around 20 genes [214]. In plants, flavonoids play various roles, for example, in the recruitment of pollinators, UV protection or protection from oxidative stress [215]. Several studies have suggested that food rich in anthocyanins provides protection against diseases such as cancer, cardiovascular disease, diabetes or obesity [214]. Most currently cultivated maize varieties produce non-pigmented yellow kernels, although the plants have often retained the ability to produce pigment in various tissues, particularly in response to biotic or abiotic stresses [214]. Those pigments are mainly anthocyanins. The flavonoid pathway in maize is regulated by bHLH and MYB transcription factors, among which well-studied paramutation loci like *r1 (red 1), b1 (booster 1), pl1 (purple 1)* and *p1 (pericarp 1)* [89,174,216].

In this thesis, we studied paramutation, a phenomenon in which communication *in trans* between alleles results in a susceptible allele gaining an epigenetic profile similar to that of the other, inducing allele. Paramutation phenomena have been described in various organisms

including maize, Arabidopsis, pea, tomato, Drosophila and mouse [81,83–87]. In Chapter 5, we studied paramutation at the *b1* locus in maize.

Paramutation: The *B'* repeat is not a classic RdDM locus

Work on recombinant maize inbred lines (RILs) reported that paramutation-like switches occurred in about 10% of the differentially methylated regions (DMRs) identified [46], suggesting that paramutation is a more common phenomenon than previously thought. In Chapter 5 of this thesis we studied paramutation at the *b1* locus in maize. The two *b1* epialleles that engage in paramutation are the paramutagenic *B'* and the paramutable *B-I* epiallele [37,91]. The *B-I* epiallele shows a 10-20 fold higher *b1* expression than *B'* [90]. This expression difference causes dark purple pigmentation in *B-I* and light purple pigmentation in *B'* plants [13], and is due to differences in DNA methylation and histone modifications at an upstream regulatory element (enhancer). *B'* exhibits high levels of DNA methylation and repressive histone modifications (H3K9me2 and H3K27me2), while *B-I* displays low DNA methylation and, when expressed, activating histone modifications (H3ac) [37,91,176]. Upon combining *B'* and *B-I* in a cross, *B-I* heritably acquires the epigenetic profile of *B'* in terms of methylation and histone modifications [90]. The mechanism of paramutation is not entirely understood. Most evidence points towards an RNA-dependent DNA Methylation (RdDM) model for paramutation. In this model small RNAs derived from the inducing (paramutagenic) allele are acting on the susceptible (paramutable) allele in the same nucleus and trigger its DNA methylation and consequent silencing [81,82]. However, the role of small RNAs in the process has never been clearly established. For the *b1* locus, it was demonstrated that siRNAs are produced from the paramutagenic *B'* repeats, but siRNAs are also produced, in seemingly similar quantities, from the paramutable *B-I* repeats and even from the

single repeat of a *b1* allele that does not engage in paramutation (neutral allele) [99]. These findings suggested that the presence of small RNAs alone is not sufficient for paramutation to occur. However, Arteaga-Vazquez *et al.* [99] used tissue for small RNA profiling from a stage in plant development (immature ears) where paramutation has long been initiated [91]. We hypothesized, that in early developmental stages (like embryogenesis) where paramutation is likely initiated, small RNAs from the *B'* repeats are present at higher levels than small RNAs from the *B-I* repeats. In line with that idea, RdDM-related factors, like MOP1 and MOP2*,* are not only highly expressed in immature ears, but also during embryogenesis in maize (Maize eFP Browers, [179]).

Therefore, in chapter 5, we used targeted bisulfite sequencing of the DNA and deep sequencing of the small RNAs in maize embryos to study the *b1* paramutation process in early plant development. Paramutation at the *b1* locus is mostly restricted to an enhancer region about 100 kb upstream of the *b1* transcription start site [175]. This enhancer consists of 7 copies of a unique 853 nucleotides (nt) sequence in a tandem hepta-repeat arrangement. Multiple copies of the repeat unit are required for enhancer activity and paramutation [37]. A tandem hepta-repeat of the 413bp 5'-half of one repeat unit (termed Fragment A, FA) is sufficient to induce paramutation and harbors enhancer function, while a tandem hepta-repeat of the 444-bp 3'-half of a repeat unit (FB) by itself does not engage in paramutation but may still harbor enhancer function [178]. For our DNA methylation analysis we focused on the FA fragment, splitting it further into two regions: the upstream FA.1 and the downstream FA.2 region.

We found that in the inactive *B'* epiallele the FA fragment was highly methylated (as expected and as shown in prior studies [91]). This high DNA methylation was, however, primarily occurring in symmetric sequence context (CG and CHG; H=A, C or T); asymmetric methylation (CHH) was

almost lacking. In accordance with the low CHH methylation level we found relatively low levels of small RNAs mapping to the hepta-repeat in *B'*, whereby the majority of the siRNAs that did map, mapped to FA.1. These findings were unexpected as methylation of cytosines in CHH context (i.e. mCHH) and relatively high small RNA coverage are considered a hallmark of the RNA-dependent DNA methylation pathway (RdDM) [17,28]. RdDM is known to be involved in the establishment of methylation in all sequence contexts and in the maintenance of CHH methylation [17], and small RNA abundance at genomic regions was shown to correlate very well with mCHH levels in maize [46]. Mutations in genes involved in the RdDM pathway (partially) released silencing of *B'* and prevented paramutation, which led to the conclusion that the *B'* repeats are an RdDM locus [92–95,176]. The paradoxically low mCHH levels we observed at the *B'* repeats (Chapter 5) point towards RdDM targeting *B'* only marginally compared to other RdDM loci. This possibility is supported by genome-wide analyses that show that RdDM loci in maize exhibit a broad range of mCHH levels including very low levels of mCHH (Figure 7, Chapter 5).

In some mutants of the RdDM pathway, transcriptional silencing of *B'* is partially released [81,82,176]. This release hardly affects the symmetric methylation profile at the hepta-repeat [74], probably due to the undisrupted activity of the maintenance DNA methylation machinery (not part of RdDM) involving ZMET1 and chromomethylases, such as ZMET2 and ZMET5 [28,176]. However, the transcriptional activation in mutants is associated with an increase in an activating histone modification (H3ac) and a decrease in repressive histone modifications (H3K9me2 and H3K27me2) [176]. Perhaps, RdDM directly influences histone modifications, or (less likely) those effects on histone modifications are a bare consequence of the transcriptional activation,

We conclude that even though the RdDM pathway is involved in paramutation at the *b1* locus, the *B'* epiallele is not an RdDM locus in the classic sense: there is little methylation of the asymmetric CHH sites, and the small RNA levels are low.

An mCHH island at the paramutable *B-I* allele may play a role in paramutation

In Chapter 5 we also describe that the FA fragment at the paramutable *B-I* hepta-repeat consists of a subregion (FA.2) that is RdDM-targeted, as illustrated by high CHH methylation (mCHH) and siRNA coverage, and an unmethylated region (FA.1). This RdDM-targeted FA.2 region fulfills the criteria of a so-called mCHH island. mCHH islands have been shown to be present at the boundary between eu- and heterochromatic (i.e. transcriptionally active and inactive, respectively) regions in maize [29]. We speculate that, together with small RNAs derived from *B'*, the CHH-methylated *B-I* repeats support the spreading of DNA methylation into the unmethylated *B-I* repeat sequences during paramutation. The identification of methylation nucleation sites adjacent to this FA.2 RdDM target region supports the spreading idea. In conclusion, our findings suggest a role for the RdDM process at the FA.2 region of the paramutable *B-I* epiallele during paramutation.

A similar observation in terms of an mCHH region at the paramutable allele that may serve as a nucleation region has been described for another paramutation locus in maize: the *p1* locus. Paramutation at the *p1* locus resembles *b1* paramutation in many aspects [81], except that the *p1* enhancer contains fragments from known transposons, among them a MULE fragment [189]. At the paramutagenic (*P1-pr*) epiallele, the enhancer is fully methylated in symmetric context (CG, CHG), as is the MULE fragment. At the paramutable (*P1-rr*) version of the epiallele, the

MULE fragment displays, beside high symmetric methylation, also CHH methylation [188]. This could be an indication for a similar role of an mCHH region in supporting the spreading of methylation at other paramutation loci.

Observations from Arabidopsis at two epialleles of the *FLOWERING WAGENINGEN* (*FWA*) locus [36] indicate that particular epigenetic profiles may be required for paramutation to occur. Upon crossing lines harboring the endogenous methylated (silent) and unmethylated (active) locus, no paramutation occurred and the epialleles kept segregating in a Mendelian fashion. Both the active and silent endogenous *FWA* loci produced small RNAs, at similar levels [217]. Strikingly, a transgenic *FWA* locus could become *de novo* methylated in the presence of the silent endogenous *FWA* locus, but not in the active *FWA* [217]. These findings support the idea that small RNA production at differentially methylated epialleles is not the only prerequisite for paramutation. Additionally, the alleles engaging in paramutation may need to exhibit particular (matching) features for the process to start. Perhaps, a methylated subregion at the paramutable allele is required to allow paramutation.

Future research should test if the FA.2 mCHH region in *B-I* is really required for the spreading of methylation during paramutation. So far the *b1* repeat unit has only been split and analyzed in two sub fragments: FA and FB [178]. Thus, it should be interesting to investigate if the entire FA fragment is required for paramutation and/or enhancing expression or only a part of it. By creating transgenes harboring a hepta-repeat of the FA.1 or FA.2 fragments one could address the relevance of the mCHH island at *B-I* for paramutation. Our hypothesis predicts that if a transgene harbors only FA.1 (i.e. no mCHH island) in a hepta-repeat organization, it will not engage in paramutation (or paramutation would at least be slowed down).

Model of Paramutation based on small RNAs and features of the paramutable allele

Many DMRs produce small RNAs but not all undergo paramutation-like processes [79,217], hence the production of small RNAs is not the only determinant of the process. We propose that particular other features of the paramutagenic and paramutable allele ultimately determine whether paramutation occurs at a locus or not.

Our paramutation model involves small RNAs derived from the paramutagenic allele and mCHH islands at the paramutable allele. We propose that i) small RNAs from the *B'* repeats act on the unmethylated regions of each of the *B-I* repeats, and that as a result ii) through spreading of DNA methylation from the mCHH islands into the unmethylated *B-I* repeat regions iii) the mCHH islands at the *B-I* hepta-repeat gradually change into regions with mostly symmetric methylation (Figure 8, Chapter 5). We hereby speculate that the mCHH islands at the *B-I* allele play a role in spreading of the methylation, and hence in paramutation of *B-I* into *B'*.

The exact molecular mechanisms underlying our observations, particularly concerning the exact role of RdDM in paramutation, remains to be addressed. Experiments could include studying paramutation at the repeats in F1 embryos in RdDM mutant backgrounds. Thereby one can investigate the effect on DNA methylation, / histone modification and gene expression during the paramutation process, caused by loss of RdDM.

Paramutation as a player in epigenomic and phenotypic diversity

Paramutation is form of gene regulation that may play a more global role in the establishment of epigenetic variation, phenotypic diversity and hybridization-linked phenomena like heterosis. In terms of evolutionary

significance, it has been speculated that paramutation might provide a mechanism for transmitting favorable expression states to progeny or that it could be involved in establishing functional homozygosity in polyploids [88]. Furthermore, more than a decade ago it has already been proposed that the mechanism may play a role in heterosis and inbreeding depression [88].

A study on recombinant maize inbred lines (RILs), derived from crosses between the maize B73 and Mo17 lines, reported paramutation-like switches in about 10% of the differentially methylated regions [46]. However, Li *et al*., comparing inheritance patterns at DMRs between near-isogenic lines (NILs) derived from crosses between the same two inbred lines (B73 and Mo17), found evidence for paramutation-like switches in only 3 % of the DMRs [141]. The discrepancies between these studies will in part be due to biases in the different methods used to examine DNA methylation, bisulfite sequencing and meDIP-chip [141]. Nonetheless, these findings suggest that paramutation-like switches occur more frequently than expected from the small number of well-characterized examples. To what extent those switches affect phenotypic diversity is not known yet.

In domesticated species that are mainly selfing, for instance crops like tomato or maize, DNA sequence variation may be more limited than in exclusively outcrossing species, and phenomena like paramutation might have more relevance for phenotypic diversity in the former species. For instance, tomato as a domesticated species, shows rather limited genetic diversity, while it retains considerable phenotypic variability [218–220]. Accordingly, epigenomic divergence may play a role in tomato varieties. One example of a paramutation case in tomato has been described at the *SULFUREA* (*SULF*) locus [221], and this phenomenon is associated with changes in DNA methylation and siRNA levels at the *SLTAB2* promoter

[222]. Plants carrying the silent *sulf* allele have a chlorotic phenotype, i.e. yellowing caused by insufficient chlorophyll production [84]; this is associated with reduced auxin [223].

One of the first steps towards understanding the relevance of epigenomic diversity and paramutation phenomena for phenotypic diversity in crop plants are comparative genome-wide studies on epigenetic modifications in crop varieties, correlating these with crop yields and other desired characteristics.

Perspectives

It has long been viewed as a general biological principle that the phenotype is the result of genetic (G) and environmental influences (E) and the interactions between them (P= G + E + G-E). The phenotypic variance should then consists of $V_P = V_G + V_E + V_{GE}$, where V refers to a sort of squared partial differential of phenotype with respect to any of the three types of cause. This principle disregarded epigenetics as a contributing factor. However, given that the contribution of epigenetics to phenotypic diversity is gaining more appreciation as genomes and their epigenetic modifications are being better characterized, the classic formulations need to be supplemented with epigenetic variation as a defined parameter.

Studying epigenomic variation in populations faces challenges different from the ones encountered when studying genomic variation. These challenges include (i) that epimutation rates can be higher than mutation rates, (ii) that paramutation-phenomena can result in non-Mendelian inheritance patterns of epialleles, and (iii) that epigenetic polymorphisms can be dependent and independent of genetic variation.

This thesis combined an extensive phenotypic epigenomic population study on heterosis in Arabidopsis epigenetic hybrids with a recently developed mapping approach for $QTLs^{epi}$ (Chapter 2 and 3). The

phenotypic data was then expanded with molecular insights into different epigenetic hybrids (Chapter 4). And finally, in chapter 5, we studied one locus in maize aiming to contribute to the mechanistic understanding of a phenomenon that occurs during hybridization (paramutation; Chapter 5).

With this thesis, I hope to have contributed to the acknowledgement that epigenetics is a phenomenon with realism and impact, and to new inroads into mechanistic understanding, supported with data from accessible experimental systems.