Phenotypic variation in plants

Roles for epigenetics

Lauss, K.

Publication date
2017
Document Version
Other version
License
Other

Chapter 1

Introduction

A phenotype refers to traits or characteristics of an organism. The variability in phenotypes among individuals of a population can be described as phenotypic variation. While DNA sequence variation is known to be a major driver of such phenotypic variation, epigenetic variation has long been disregarded as a contributing factor. Epigenetics encompasses modifications of the DNA sequence and chromatin that can affect gene activity. By now, suitable tools have been developed to better investigate the contribution of epigenetic variation to phenotypic divergence.

In this thesis we show that epigenetic variation contributes to an interesting phenomenon occurring in hybrids: heterosis. Heterosis refers to superior performance of a hybrid compared to its parent varieties. We describe and apply a recently developed mapping approach for epigenetic quantitative trait loci (QTL^{epi} mapping) for identifying epigenetically divergent regions potentially underlying some of the heterotic traits we observed. Moreover, we perform a detailed one-locus study on a phenomenon that can occur during hybridization at regions displaying different epigenetic profiles: paramutation. Paramutation refers to communication between epigenetically divergent regions resulting in one region acquiring the epigenetic profile of the other. Our data indicate that characteristics of both participating regions are important for paramutation.

In sum, our observations contribute to an understanding of the complex relationship between epigenetic and phenotypic variation.

Genotype – phenotype – environment

The first discrimination between “heritable dispositions” of organisms (genotype) and the physical characteristics affected by to those heritable dispositions (=phenotype) has been made in 1909 [1]. Since then, more refined definitions of phenotype have been proposed for example: “A

plant phenotype is the set of structural, physiological and performance-related traits of a genotype in a given environment" [2].

Phenotypic traits can be categorized into qualitative and quantitative traits. Qualitative traits have distinctly defined phenotypic categories, are often regulated by one (monogenic) or a few genes and the influence of the environment on the trait is rather subtle (e.g. resistance to particular pathogens) [3]. Quantitative traits, on the other hand, occur in a continuous distribution of phenotypic values. These distributions in phenotypic values are caused by the influence of the environment and by the fact that usually several loci (polygenic) are involved in forming a quantitative trait [3,4]. One useful way of studying the effect of a gene on a phenotype is by silencing or knocking-out the gene of interest. However, redundancy of genes or pathways can lead to mutants not showing a phenotype [3]. Also, particular phenotypes may only show under certain circumstances or in particular genetic backgrounds [3]. An alternative to mutant-screens is the exploitation of natural variation among populations. As natural trait variation is often quantitative, in addition to mutant screens, other tools are needed to identify the underlying genes [3]. To that end, statistical methods in combination with molecular marker technologies have been developed to map quantitative trait loci (QTLs) in genomes allowing exploring natural variation across subspecies or accessions [3,4].

Exploring phenotypic variation by genetic mapping approaches

Natural variation within species has been effectively exploited to associate genotypic divergence with phenotypic trait variation and has been recognized as a valuable source for agricultural crop improvement [3]. In order to study genotype-phenotype relationships, different parental lines with varying genotype and phenotype have been used to create bi-parental mapping populations. Hereby, Arabidopsis has proven a useful

resource with its short life cycle and small genome. In principle, any type of experimental segregating population can be used for mapping approaches such as F2 populations or progeny of backcrosses of F1 to a parental line [3,4]. For Arabidopsis plants recombinant inbred lines (RILs) are frequently used for mapping approaches. RILs are produced by rounds of repeated backcrossing or self-fertilization (selfing) [5]. RILs have the advantage over an F2 population that they can be replicated endlessly (immortal mapping population), and that genotyping is only necessary once, while phenotyping can be performed numerous times, for various traits and in various conditions [4].

"Above" Genetics

DNA, the genetic information of a cell, is packaged together with histones into chromatin [6]. Chemical modifications at DNA bases and histones, and the incorporation of histone variants, constitute an additional layer of information referred to as epigenetics [6,7].

The term epigenetics was introduced by Conrad Waddington in 1942 who defined it as " the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" [8] or in simpler words the way of one genotype giving rise to various phenotypes (i.e. different cell types of an organism). Together with a substantial increase in knowledge in this field of research, also the definition of "epigenetics" has evolved. Several follow-up definitions involved an emphasis on the heritability of epigenetic changes [9,10] but there are also more recent definitions advocating less stringency in terms of heritability in order to include stable modifications, but also short-lived modifications, and modifications occurring in non-dividing cells (e.g. neurons) [11,12]. One of those definitions describes epigenetics as "the study of any potentially stable and, ideally, heritable change in gene

expression or cellular phenotype that occurs without changes in Watson-Crick base-pairing of DNA" [12].

Epigenetic mechanisms in plants

The major epigenetic mechanisms in eukaryotes involve activating and repressing histone modifications, DNA methylation at cytosine-bases in the DNA sequence (including the small interfering RNAs capable of inducing DNA methylation), and the incorporation of histone variants into chromatin [13,14]. These mechanisms are involved in the establishment and maintenance of repressed and active states of genomic regions. Here we will introduce the first three mechanisms.

Histone Modifications

Histone modifications, like acetylation and methylation, are implicated in particular chromatin states [6]. Acetylation of histones, for example of histone 3 (H3ac), increases accessibility of the DNA sequence in the affected regions and is associated with an active chromatin state [6]. Methylation of histones is associated either with an active or repressed chromatin state depending on the position of the modification on the histone tails [6]. Tri-methylation of histone 3 at lysine 27 (H3K27me3) and H3K9me2 are associated with repressed genes and heterochromatin, respectively, while H3K4me2 is associated with active genes [6]. Repressive histone modifications (H3K9me2) are frequently interconnected with DNA methylation [15,16].

DNA methylation and small interfering RNAs

A very well-studied epigenetic modification that is associated with specific transcriptional expression states is DNA methylation of cytosine (C) residues [7]. In plants C's in all sequence contexts can be methylated (CG, CHG, and CHH, where H represents non-G nucleotides) [7]. A pathway

termed RNA-dependent DNA methylation (RdDM) is involved in mediating *de novo* methylation in all sequence contexts via small interfering RNAs [17]. The canonical RdDM pathway in Arabidopsis is described as follows: DNA-dependent RNA polymerase IV (Pol IV) produces transcripts that are converted into double-stranded RNA by RNA-dependent polymerase 2 (RDR2) [17]. DICER-LIKE 3 (DCL3) processes this double-stranded RNA into 24nt small interfering RNAs (siRNAs) [17]. The siRNAs are loaded into ARGONAUTE 4 (AGO4) and the complex binds to its complementary transcript produced by RNA-Polymerase V (Pol V) [17]. This results in the recruitment of DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM 2), which mediates DNA methylation in all sequence contexts [17].

Once established, symmetric methylation (mCG and mCHG) can be maintained by maintenance enzymes METHYLTRANSFERASE 1 (MET1) and CHROMOMETHYLASE 3 (CMT3), even through meiotic divisions [17]. Asymmetric methylation (mCHH) on the other hand relies on RdDM for both, its establishment and maintenance via the enzyme DRM2 [17]. CHH methylation is also mediated by another (siRNA-independent) pathway involving histone dimethylation (H3K9me2) and CHROMOMETHYLASE 2 (CMT2), which acts mainly on pericentromeric heterochromatin [18]. In addition, the chromatin remodeler DECREASE IN DNA METHYLATION 1 (DDM1) plays a central role in mediating DNA methylation at percentromeric heterochromatin by allowing methyltransferases to access the DNA [17].

DNA methylation in plants is found at genes but more frequently at transposons [19]. Loss of methylation can cause transposon re-activation, emphasizing the relevance of DNA methylation for transposon silencing [18,20,21]. DNA methylation at genes can be present in both promoter regions and gene bodies [22,23]. Promoter methylation commonly results in reduced gene transcription while gene-body methylation (mainly CG) is associated with moderate-to-high gene expression [22,23].

While DNA methylation is low or even absent in some eukaryotes like Drosophila or particular yeast species, in mammals defects in DNA methylation are often embryonic lethal [7,24] emphasizing the importance of DNA methylation. The effect of DNA methylation defects in plants depends on the species: in Arabidopsis DNA methylation mutants do sometimes show phenotypic defects but can still be propagated [25–27], by contrast, similar mutants in maize are mostly embryonic lethal [28]. It has been proposed that the reason for these differences lie within different genome organizations: Arabidopsis has a small genome with around 36% transposons (TEs) that are mostly localized in pericentromeric regions, while maize has a large genome with about 85% consisting of transposons, which are more dispersed throughout the genome and often located close to genes [29–32]. There are also differences concerning pathways involved in TE-silencing like RdDM: In Arabidopsis, RdDM targets mainly short transposons, repeats and transposon edges [18] and a siRNA-independent pathway involving CMT2 acts mainly on long TEs in the pericentromeric regions. In maize it has been shown that RdDM-dependent mCHH occurs close to genes or conserved non-coding sequences [29,33]. Such regions, termed mCHH islands, were proposed to enforce boundaries between eu- and heterochromatin [29,33]. No CMT2 homologs have been found so far [18]. There is evidence that CMT2 function could be covered by other chromomethylases (*Zmet2* and *Zmet5*) [28], but it may also be that maize lacks the CMT2 function.

Epigenetic variation across plant subspecies

There are various examples of the involvement of epigenetic variation in phenotypic variation in plants [34,35]. For instance, a number of epialleles that have been described in different species, including Arabidopsis, maize, rice and tomato [36–41]. Epialleles are allele variants that have the

same DNA sequence but differ in their epigenetic modifications, thereby causing differences in gene expression.

The epigenetic variation occurring among plant subspecies harbors a certain potential for undiscovered epialleles. Patterns of DNA methylation are shown to vary extensively among natural Arabidopsis populations and between subspecies of rice and maize [19,42–44]. Similarly small RNA differences have been found among subspecies of maize [42,45,46] and Arabidopsis [47–49] and variation in histone modifications has been reported in studies comparing Arabidopsis accessions [50,51].

Exploring epigenetic variation

A powerful resource for studying and quantifying epigenetic variation and its association with plant traits are Arabidopsis epigenetic recombinant inbred lines (epiRILs) [52,53]. epiRIL populations allow examining the effects of epigenetic variation largely independent of genetic variation. The two existing epiRIL populations are near isogenic in their DNA sequence but are a mosaic in terms of their epigenetic profiles [52,53]. These two populations have been created by crossing Columbia (Col-0) wildtype with Col-0 lines carrying mutations in *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 asymmetric methylation and particular histone modifications [25,54]. Loss of DDM1 results in a ~ 70% reduction in DNA methylation [26] at transposable elements but also affects genic loci by reducing CG methylation and causing CHG hypermethylation in gene bodies [18]. The available epiRILs show phenotypic variation in various traits, confirming the concept of DNA methylation affecting quantitative traits [52,53,55]. To date, epiRILs have been used to determine the contribution of epigenetic variation to particular traits [56], phenotypic

plasticity in response to stress [57] and heterosis [58]. Moreover, analogous to RILs being used to identify causal genomic loci explaining trait variation by mapping quantitative trait loci (QTLs), *ddm1*-epiRILs were used to detect causal loci in the epigenome ($QTLs^{epi}$). Thereby, $QTLs^{epi}$ were identified for flowering time, root length, and for phenotypic plasticity in response to stress [56,57].

Future research and breeding efforts may aim towards studying and potentially exploiting epigenetic variation also in crop plants. Breeding strategies in the crop plant canola already resulted in epiLines that show improved energy use efficiency and drought resistance compared to their isogenic counterparts [59,60].

Hybridization and Heterosis

Hybrids are the progeny of two parental lines that differ in their (epi)genomes. Thereby a unique (epi)genetic background that is usually not just the sum of its parental contributions is created. In nature, plant hybridization occurs naturally, for example by insect pollination or wind pollination, and particular hybridization events have been implicated in plant speciation [61].

There are two possible ways of hybridization, one occurs between different species (interspecific) and the other amongst different varieties of the same species (intraspecific) [62,63]. Interspecific hybridization encompasses the formation of allopolyploids, which are hybrids with multiplied sets of chromosomes derived from different species. Many crops, like cotton or bread wheat, are allopolyploids [64,65].

Certain rice and maize varieties are grown as intraspecies hybrids [62].

The novel combination of parental (epi)genomes in hybrids can result in the occurrence of certain phenomena. For instance, hybridization can be accompanied by incompatibilities, such as hybrid sterility or lethality [66]. Causes for sterility can, for example, be DNA sequence divergence at

particular genomic loci, difficulties in chromosome pairing (in case of interspecies hybrids) or failure of epigenetic silencing [66].

Heterosis (hybrid vigor) is another phenomenon resulting from hybridization [63,67]. It refers to hybrid progeny outperforming its parents in particular traits. The phenomenon is very important for the agricultural industry, as it has increased crop yields tremendously over the past decades. In the aim of understanding why outcrossing species are prevalent in nature, Charles Darwin was among the first to document beneficial effects of outcrossing versus harmful effects of inbreeding in plant species [68]. In the early twentieth century, studies in maize reported that hybrids often show a substantial degree of heterosis compared to their inbred parents, leading to the implementation of breeding programs exploiting heterosis [69–71].

Since then, various efforts in quantitative genetics and molecular approaches have aimed to understand hybrid vigor [63,67,72]. Early studies proposed mainly genetic models to explain heterosis: the classical models are the dominance and overdominance hypotheses. In short, the dominance hypothesis claims that heterosis arises from a complementation of deleterious alleles [69,73,74], while the overdominance hypothesis postulates that heterozygosity at particular alleles is advantageous over homozygosity [69,70]. Later, other genetic models have been proposed, including epistasis, which refers to non-allelic gene interactions contributing to heterosis [75]. There are lines of evidence and examples for each hypothesis, suggesting that multiple mechanisms may apply or that mechanisms underlying heterosis may differ on a case-by-case basis. In addition, it has been shown that in heterotic hybrids, epigenetic changes relative to the parental lines occur concerning levels of small regulatory RNAs and/or DNA methylation for Arabidopsis, maize, rice and tomato, indicating a contribution of epigenetic mechanisms to heterosis [45,47,76,77]. To address the

contribution of epigenetics to heterosis independently of genetic variation, lines from two populations of epigenetic recombinant inbred lines (epiRILs) [52,53] were used to study heterosis in epigenetic F1 hybrids (epiHybrids) [58]. In one epihybrid, generated by crossing a *met1*-derived epiRIL with Col-0 wildtype, 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]. On the molecular level, evidence for an "epigenomic shock" has been observed in a non-heterotic epihybrid (derived from Col-0 and an isogenic *met1*-mutant line) [78]. This shock involved the formation of novel epialleles and activation of transposons which were not active in either parent [78]. The described epigenomic shock also involved an interesting phenomenon that has been observed in genetic hybrids between Arabidopsis accessions: *in trans* communication between alleles can result in a region acquiring a similar epigenetic profile as the allelic region on the homologous chromosome [79]. When such changes in epigenetic profiles are transgenerationally stable, they are termed paramutation [80].

An epigenetic phenomenon resulting in non-Mendelian inheritance patterns: Paramutation

Paramutation is an epigenetic phenomenon that refers to *in trans* communication between alleles with different epigenetic modifications (often including DNA methylation) resulting in the susceptible allele acquiring the same (or a similar) epigenetic profile as the inducing allele [81,82]. Paramutation phenomena have been reported in a variety of organisms including maize, Arabidopsis, pea, tomato, Drosophila and mouse [81,83–87]. Most described well-studied paramutation cases have been discovered because they led to an easily detectable (unexpected) phenotype [88]. However, advances in genome-wide analyses of

epigenetic marks permit the detection of paramutation or paramutation-like cases that do not involve a visible phenotype and first reports already revealed widespread paramutation-like switches in maize [46]. Alleles inducing paramutation are termed “paramutagenic” while alleles susceptible to paramutation are termed “paramutable”. Alleles that are neither paramutagenic nor paramutable are termed neutral alleles.

In maize, paramutation has been well studied at several loci including *red1* (*r1)*, *pericarp 1 (p1)*, *purple 1 (pl1)* and *booster 1 (b1)*, which all encode transcription factors involved in plant pigmentation [89]. The focus of this thesis is the *b1* locus, which encodes a transcriptional regulator of the anthocyanin pigmentation pathway [90]. The two *b1* alleles that engage in paramutation, the paramutagenic *B’* and the paramutable *B-I*, are epialleles, meaning they share the same DNA sequence but vary in their epigenetic profile [37,91].

The mechanism behind paramutation is not well understood yet. What has been demonstrated is, that a number of proteins orthologous to components in the RdDM pathway of Arabidopsis are necessary for paramutation in maize [82]. These RdDM proteins are termed *mediators of paramutation* (*mop*) or *required to maintain repression* (*rmr*) genes [82] and include MOP1 (ortholog of RDR2), MOP2/RMR7 (ortholog of NRP(D/E) 2a -second largest subunit of Pol IV and V) and MOP3/RMR6 (ortholog of NRPD1 – the largest subunit of Pol IV) [92–95]. As expected for factors involved in RdDM, mutations in *Mop1*, *Mop2* and *Mop3*/*Rmr6* have been shown to cause depletion of siRNAs, and reduction of DNA methylation in particular transposon families [28,93,94,96–98] and in regions with CHH methylation [28,98].

Current models for paramutation are based on an RdDM-like mechanism, whereby small RNAs derived from the inducing (paramutagenic) allele are acting on the susceptible (paramutable) allele and trigger its DNA methylation and silencing [81,82]. However, despite several efforts in

maize [99,100], the role of siRNAs in the process has never been clearly established. For the *b1* locus it was demonstrated that siRNAs are produced from the *B'* repeats, but siRNAs were also produced, in seemingly similar quantities, from the *B-I* repeats and even from the single repeat of a neutral *b1* allele [99]. This indicates that the presence of small RNAs alone is not sufficient to make an allele paramutagenic.
Paramutation is a very interesting form of gene regulation that could have a more global role in the establishment of epigenetic variation and hybridization-linked phenomena like heterosis than initially anticipated.

Scope of this thesis

In this thesis we aim to contribute to an understanding of the complex relationship between epigenome and phenotype in two main projects. In the first one, we investigated the contribution of epigenetic modifications to heterosis in Arabidopsis epiHybrids by combining a phenotypic study with mapping of epigenetic quantitative trait loci and analysis of DNA methylation and gene expression. In the second project, we investigated the mechanisms underlying paramutation by focusing on the *b1* locus of maize. At the *b1* locus, the *B-I* epiallele is undergoing paramutation when combined with the *B'* epiallele during hybridization, thereby creating a non-Mendelian inheritance of an epigenetic profile.

Chapter 2 is a methodical chapter that guides through the recently developed method of epigenetic QTL mapping in Arabidopsis. We describe aspects of the creation of epigenetic recombinant inbred lines, phenotyping, data analysis and confirmation of candidate regions.

In **chapter 3** we aimed to investigate the role of epigenetic divergence in heterosis. To that end we used a selection of epiRILs to create epigenetic F1 hybrids (epiHybrids), which we screened for heterotic phenotypes in a

range of traits. We found both positive and negative heterotic phenotypes among the epiHybrids. Moreover, by implementing a QTL^{epi} mapping approach we could associate particular DNA methylation profiles in the parental lines with some of the heterotic phenotypes observed. Our findings indicate that epigenetic divergence can be sufficient to cause heterotic phenotypes in various traits and that $QTLs^{epi}$ can partially explain our observations.

Chapter 4 describes molecular insights into four selected epiHybrids and their parental lines from chapter 3. We performed genome-wide bisulfite and transcriptome sequencing and describe our observations globally and at the putative QTL^{epi} regions. We describe non-additive changes in RNA expression, particularly transposons, in the epiHybrid and we report strong indications for *trans*-chromosomal (de)methylation events occurring in the epiHybrids.

In **chapter 5** we aim to shed light onto the early steps of the paramutation process at the *b1* locus in maize by performing DNA methylation profiling and small RNA sequencing. Interestingly, we found an RdDM-targeted region at the paramutable allele of the *b1* locus that likely plays a role in paramutation. Our results indicate that features of both the paramutagenic and the paramutable allele are relevant for paramutation at the *b1* locus.

Chapter 6 discusses the work presented in this thesis.