

REVIEW

Open Access



Methylome evolution in plants

Amaryllis Vidalis^{1†}, Daniel Živković^{2†}, René Wardenaar³, David Roquis¹, Aurélien Tellier^{2*} and Frank Johannes^{1,4*}

Abstract

Despite major progress in dissecting the molecular pathways that control DNA methylation patterns in plants, little is known about the mechanisms that shape plant methylomes over evolutionary time. Drawing on recent intra- and interspecific epigenomic studies, we show that methylome evolution over long timescales is largely a byproduct of genomic changes. By contrast, methylome evolution over short timescales appears to be driven mainly by spontaneous epimutational events. We argue that novel methods based on analyses of the methylation site frequency spectrum (mSFS) of natural populations can provide deeper insights into the evolutionary forces that act at each timescale.

Introduction

Cytosine methylation is a heritable epigenetic modification and a pervasive feature of most plant genomes [1–4]. It has important roles in regulating the expression of transposable elements (TEs), repeat sequences, and genes. Extensive experimental work has shown that changes in DNA methylation are associated with plant phenotypes [5–20], genome stability [21–25], polyploidization [26], recombination [27–31], and heterosis [32–40], and that such changes actively mediate environmental signaling [41–43], pathogen responses [44–46], and priming [47–49]. For these reasons, DNA methylation has emerged as a potentially important factor in plant evolution [50–53] and as a possible molecular target for the improvement of commercial crops [54, 55].

In the model plant *Arabidopsis thaliana*, cytosine methylation occurs in symmetrical CG and CHG contexts, as well as in asymmetrical CHH sequence contexts (where H = A, T, C) [56]. Extensive forward genetic

screens in this species have made tremendous progress in dissecting the genetic pathways that establish and maintain context-specific methylation patterns throughout the genome [57]. These efforts have been facilitated by parallel technological developments in measuring methylomes at high resolution [58], which have permitted detailed assessments of the molecular impact of specific mutant genotypes.

Early methylome sequencing studies of the *A. thaliana* Columbia reference accession revealed that this model plant methylates about 10.5% of its cytosines globally (30% in context CG, 14% in CHG, and 6% in CHH, approximately), maintains dense methylation within TE and repeat sequences (at CG, CHG, and CHH sites), and (on average) intermediate methylation levels in gene bodies (mainly at CG sites) [59–62]. Insights into the evolutionary origin of these methylome features and into the forces that have shaped them over time cannot be readily obtained from experimental molecular studies, but require comprehensive inter- and intraspecific comparative epigenomic analyses. A major goal of these comparative approaches is to answer the following questions: ‘What are the factors that generate inter-individual variation in DNA methylation?’ and ‘How do evolutionary forces, such as selection, recombination and drift, act on this variation?’ A recent surge in fully sequenced plant genomes and methylomes is now providing the raw material that can be used to begin to answer these questions.

To date, the methylomes of about 90 diverse plant species have been analyzed by whole-genome bisulfite sequencing (WGBS-seq) [4, 57, 63–67] or by high-performance liquid chromatography (HPLC) [68]. These species include representatives of major taxonomic groups such as angiosperms (flowering plants), gymnosperms, ferns, and non-vascular plants, which diverged nearly 500 million years ago and thus cover much of the phylogenetic breadth of the plant kingdom. (For a list of plant species whose methylomes have been analyzed by WGBS-seq or by HPLC, and are analyzed in this Review see Additional file 1.) In addition to these interspecific data, deep genome and methylome sequencing has been performed for over 1000 natural *A. thaliana* accessions

* Correspondence: tellier@wzw.tum.de; frank@johanneslab.org

[†]Equal contributors

²Population Genetics, Technical University of Munich, Liesel-Beckman-Str. 2, 85354 Freising, Germany

¹Population Epigenetics and Epigenomics, Technical University of Munich, Liesel-Beckman-Str. 2, 85354 Freising, Germany

Full list of author information is available at the end of the article



from all over the world [63, 69–75], as well as for several experimentally derived populations in *A. thaliana*, *Zea mays* and *Glycine max* [16, 17, 19, 76–80].

Here, we illustrate how these studies are beginning to provide deeper insights into methylome evolution in plants. Our review shows that long-term methylome evolution appears to be mainly a byproduct of genomic changes, such as the differential expansion of TE and repeat sequences as well as genetic mutations in pathways that control DNA methylation or transcriptional states. By contrast, short-term methylome evolution seems to be strongly dominated by heritable stochastic changes in DNA methylation (i.e., epimutations) that occur at relatively high rates and are largely independent of genomic backgrounds.

Because these two processes operate at different timescales, an obvious empirical goal is to be able to delineate their relative contributions to inter- and intraspecific methylome diversity patterns. We provide a proof-of-principle demonstration in *A. thaliana* showing that a formal analysis of the species' methylation site frequency spectrum (mSFS) in terms of epimutational processes provides a powerful framework for addressing this challenge. We argue that further applications of such modeling approaches, in conjunction with high-throughput sequencing data, will be necessary to understand the forces that shape the evolution of plant methylomes over timescales that are of agricultural and evolutionary relevance.

Methylome evolution over long timescales

Our understanding of the genome-wide properties of DNA methylation in plants has been deeply shaped by observations of *A. thaliana*, but this model plant of the *Brassicaceae* family has an unusually small and compact genome and a plastic methylome. Early comparisons between *A. thaliana* and several commercial crops, such as *Z. mays* and *Oryza sativa*, already signaled that many features of the *A. thaliana* methylome are not entirely representative of all plant species [64, 81–83]. In order to grasp the full evolutionary significance of these differences, and to be able to identify factors that can account for them, a more extensive phylogenetic sampling of plant methylomes is necessary.

Genome size and methylome diversity

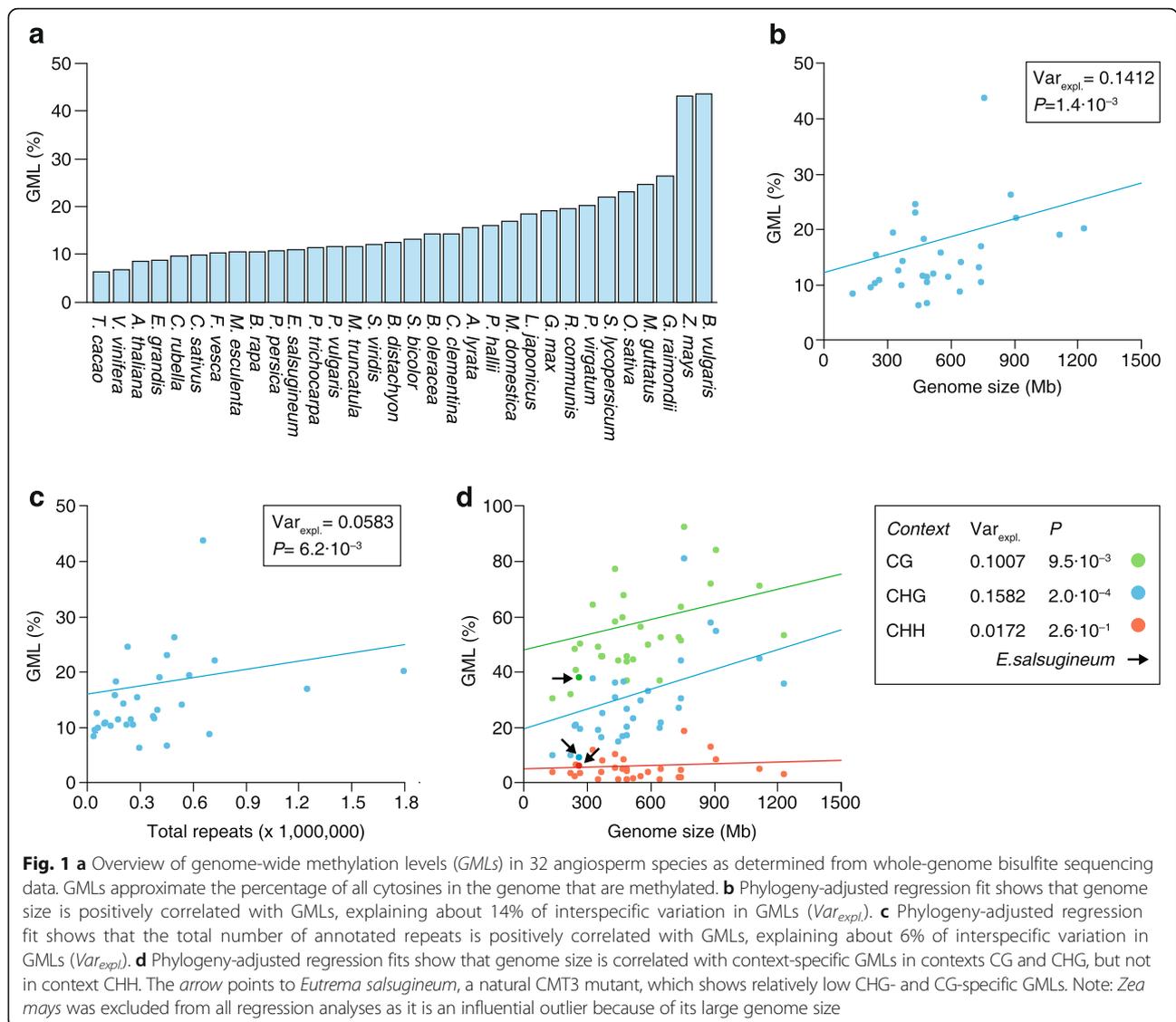
Recent comparisons of 34 angiosperm methylomes show that genome-wide methylation levels (GMLs; a measure of the percentage of all cytosines that are methylated) can vary substantially between species even within the same taxon (Fig. 1a; see Additional file 2: Figure S1 for GMLs measured by HPLC and WGBS-seq). They range from as low as 5% in *Theobroma cacao* to as high as 43% in *Beta vulgaris*, with a mean of about 16% [3, 68]. While multiple factors probably contribute to these

differences, interspecific variation in genome size is a strong predictor ([3, 68]; see Fig. 1b). Phylogeny-adjusted estimates show that about 14% of the diversity in GMLs is accounted for by variation in genome size (Fig. 1b). For every additional 100 Mbs of genomic sequence, GMLs increase by about 1.07%. This positive relationship can be explained by the fact that genome size differences are, to a large extent, the outcome of differential expansion of TEs and repeats [84, 85] (see Additional file 2: Figure S2), which are typically heavily methylated. Indeed, if the total number of annotated repeat copies in each species is used as a proxy for genome size, similar associations are detectable (Fig. 1c), although the effect sizes are somewhat smaller possibly owing to variation in repeat annotation quality [3].

These quantitative estimates support previous observations from a comparative analysis of three *Brassicaceae* species—*A. thaliana*, *Capsella rubella* and *Arabidopsis lyrata* [65]—which showed that methylome differences are mainly associated with centromeric expansion and deletion of repetitive sequences and TEs. In particular, the loss of three centromeres in *A. thaliana* relative to *A. lyrata* and *C. rubella* has led to a 10% reduction in its genome size and has a measurable impact on cytosine methylation distribution.

The extent of interspecific diversity in GMLs depends strongly on cytosine context. GMLs calculated from CG sites (i.e., CG-GMLs) vary only threefold between species, whereas for CHG-GMLs and CHH-GMLs, there is ninefold and 16-fold variation, respectively. Moreover, although genome size is associated with CG-GMLs and CHG-GMLs, there is no detectable association with CHH-GMLs (Fig. 1d). The biological source of these differences is not entirely clear, but may be at least in part due to technical difficulties in ascertaining CHH methylation states from WGBS-seq data in general [3, 4].

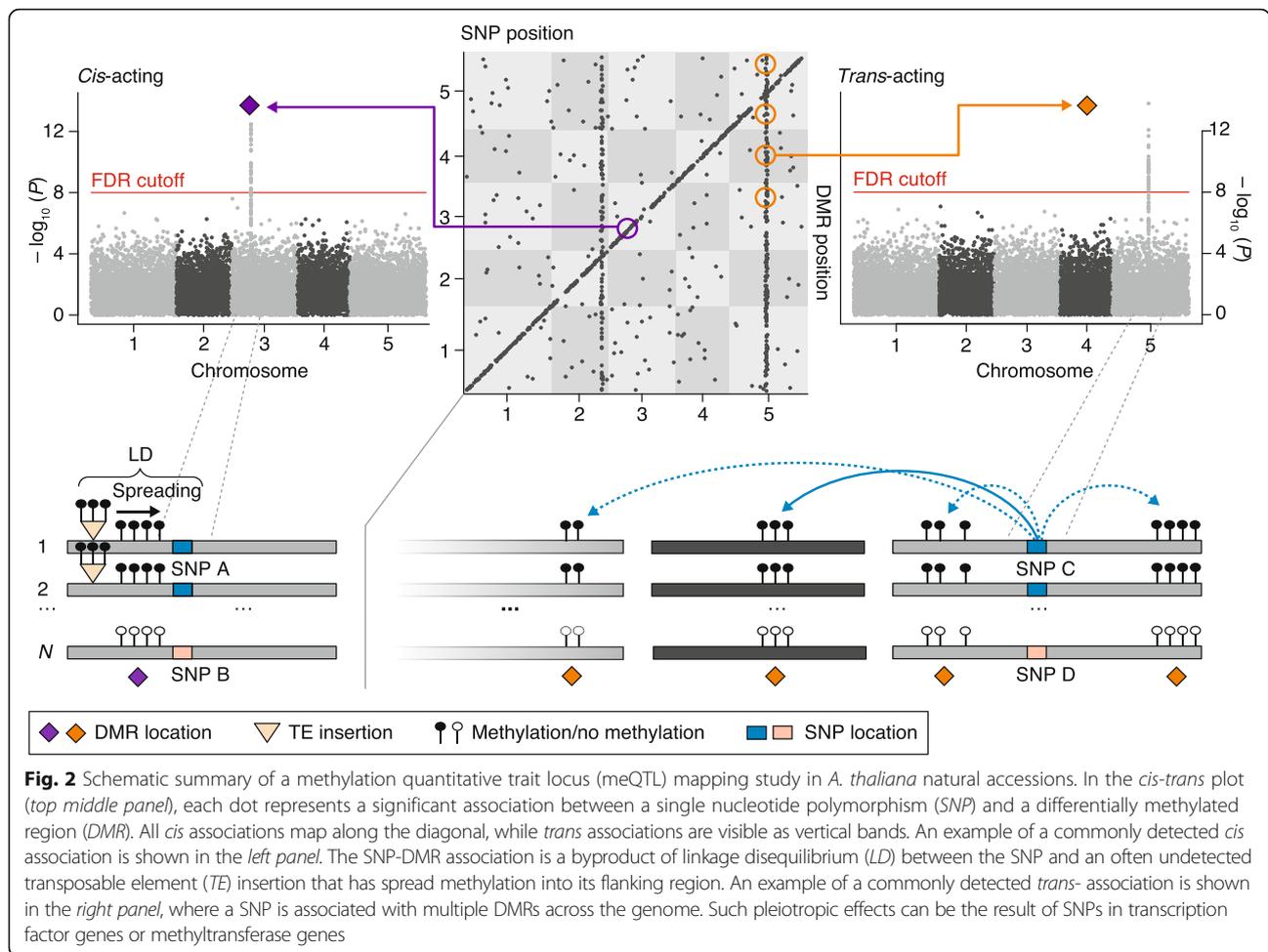
Plant genome-size evolution can be relatively rapid [85, 86]. Even closely related local populations of *A. thaliana* natural accessions differ greatly in genome length [71]. Many of these differences appear to be driven by the differential expansion of 45S rDNA copies [71], which are typically targeted by DNA methylation [87]. Considerable copy-number differences in various TE families have also been documented among worldwide samples of *A. thaliana* [69, 88, 89]. Recent methylome analyses of these samples indicate that both old and new TE copies tend to be silenced by DNA methylation [88, 89], although de novo silencing of some classes of mobile copies may require several generations and depend on copy-number frequency [90]. It is well-known that, as a byproduct of TE and repeat silencing, DNA methylation often spreads from its target sites into flanking sequences [91, 92] and produces differentially methylated regions (DMRs) at the species level (Fig. 2). In the



case of evolutionarily old insertions, such spreading-derived DMRs are effectively tagged by single nucleotide polymorphisms (SNPs) in linkage disequilibrium (LD) with the insertion sites (Fig. 2), and therefore appear as *cis* methylation quantitative trait loci (meQTL) in genome-wide association or linkage scans [63, 79, 93, 94]. Current estimates in *A. thaliana* and *Z. mays* suggest that about 20% and 50%, respectively, of all *cis*-meQTL are attributable to flanking structural variants [63, 94]. However, many TE insertions appear to have originated from very recent mobilization events and are therefore not associated with the underlying SNP haplotypes. Spreading of DNA methylation from such recent insertions produces rare epialleles that are not captured in meQTL studies, although they do contribute to methylome diversity at the species level [89].

DNA methylation pathways and methylome diversity

Beyond genome-size evolution, inter- and intraspecific diversity in genome-wide and context-specific methylation levels can also be the outcome of genetic divergence in pathways that target DNA methylation. Studies with experimental mutants in *A. thaliana*, *Z. mays* and *O. sativa* show clearly that perturbations of de novo and maintenance methylation genes can strongly affect GMLs as well as context-specific methylation patterns [19, 95, 96]. Few natural experiments exist that would permit a comprehensive evaluation of the impact of such perturbations in the wild. Recently, Bewick et al. [97] reported that two angiosperm species, *Eutrema salsugineum* and *Conringia planisiliqua*, have independently lost CHROMOMETHYLASE 3 (CMT3), an essential methyltransferase that catalyzes histone H3 lysine 9 di-



methylation (H3K9me2)-associated CHG methylation [98]. These natural mutants show significantly reduced gene body methylation as well as a reduction in global CHG methylation levels ([3, 97]; Fig. 1d).

Even single point mutations in otherwise highly homologous genes are sufficient to generate extensive methylation diversity. Dubin et al. [73], for instance, used a meQTL mapping approach to show that two *trans*-acting SNPs in the gene encoding CHROMOMETHYLASE 2 (a homologue of CMT3) substantially alter CHH methylation levels among *A. thaliana* accessions sampled from the north and south of Sweden, and another causative polymorphism in this gene has been identified in larger Eurasian samples [99]. Furthermore, Quadrana et al. [88] recently performed a genome-wide association (GWA) analysis in *A. thaliana* accessions in which they treated TE copy number as a quantitative trait. Their scan identified a candidate causal SNP in the gene encoding MET2a, a poorly characterized homologue of the CG methyltransferase MET1 [100, 101]. This example illustrates that genetic mutations that affect DNA methylation pathways can act as facilitators of genomic

changes, and set into motion complex co-evolutionary dynamics between genomes and epigenomes.

The systematic identification of similar pathway mutations is far more challenging in the context of interspecific analysis. Many genes are involved in DNA methylation pathways [56, 102], and so a comprehensive investigation of gene family phylogenies would be necessary to reveal connections between the functional conservation of specific orthologs and methylome diversity patterns. To date, such information is on the way for the CMT gene family [103], but only limited results are currently available for other methyltransferase genes or other DNA methylation-related genes [1, 4, 102, 104]. Potential insights from such an analysis are further complicated by the fact that the functional consequences of pathway mutations can be highly dependent on genomic backgrounds. This is exemplified by failed attempts to construct composite loss-of-function mutations in DNA methylation genes in *Z. mays* [19], even though similar mutations are fully viable in *A. thaliana* [95].

Nonetheless, Niederhuth et al. [3] recently argued that an indirect approach to assessing the differential

efficiency of DNA methylation pathways across species is to compare measures of intragenomic variability in site-specific methylation levels or in the degree of strand-symmetrical methylation at CG and CHG sites. In this formulation, a methylation pathway is considered robust if intragenomic variability is low and symmetrical methylation at CG and CHG is high. The fact that angiosperms differ substantially along these metrics suggests that methylation maintenance efficiency is species-dependent, even if the underlying pathway perturbations remain unknown. These metrics are certainly interesting but need to be evaluated carefully with regards to technical confounders such as mappability, genome coverage, and differential heterogeneity of the sampled tissues.

Gene-body methylation (gbM) as a neutral byproduct of transcription

Arguably one of the most enigmatic features of plant methylomes is the methylation of gene bodies. Body methylated (BM) genes have been heuristically defined as genes that methylate more than 90% of their CG sites and less than 5% of their CHG and CHH sites [105]. The latter requirement filters out genes that feature TE-like methylation patterns, perhaps because they were originally derived from TEs or contain intact or degenerate TE copies. In *A. thaliana*, about 18% of genes are BM whereas about 65% are unmethylated (UM). Unlike its repressive role in TEs and repeats, methylation in gene bodies tends to occur in moderate to highly expressed genes [62, 97]. The molecular mechanisms by which gene-body methylation (gbM) contributes to transcription, if at all, and its evolutionary significance are not fully understood.

gbM is associated with evolutionarily important genes

Indirect evidence that gbM may be evolutionarily important has come from the observation that BM and UM genes in *A. thaliana* differ markedly in sequence composition. BM genes are about twofold longer and have greater exon content [105]. Moreover, comparisons of *A. thaliana* and *A. lyrata* orthologs show that the ratio of nonsynonymous (K_A) to synonymous (K_S) substitutions is markedly lower in BM than in UM genes ($K_A/K_S = 0.198$ and $K_A/K_S = 0.23$, respectively; $p < 10^{-5}$), suggesting that BM genes are subject to stronger evolutionary constraints. Interestingly, in addition to the lower K_A/K_S ratio, BM genes seem to evolve at a slower rate in general. This is evidenced by the fact that the actual rate of, presumably neutral, synonymous (K_S) and intron (K_{INT}) divergence is significantly reduced in BM compared with UM genes ($K_S = 0.122$ in BM and 0.140 in UM, $K_{INT} = 0.107$ in BM and 0.137 in UM). In support of this argument, Takuno and Gaut [105] showed that nucleosome occupancy is positively correlated with

K_S and K_{INT} values, attributing this to more efficient DNA-repair machinery in nucleosome-free regions [105]. However, the DNA-repair argument does not readily extend to CG dinucleotides: BM genes are highly depleted in GC content as well as in the proportion of CpG dinucleotides compared with UM genes, which reflects the well-known mutagenic potential of methylated cytosine to change to thymine as a result of deamination [105]. That this difference in CG content is so visible in current sequencing data suggests that methylation levels in gene bodies must have been maintained for significant evolutionary periods.

The selection hypothesis

But how can gbM be maintained as evolution proceeds while methylated cytosines are continually lost through deamination? One explanation for this paradox is that gbM, itself, is under positive selection, which would result in an equilibrium CG content that is defined by the balance between the rate of deamination and the strength of selection [106, 107]. This selection hypothesis implicitly assumes that gbM is essential for gene function, and should therefore be conserved between orthologs across plant species. Initial methylome comparisons between two related grasses, *Brachypodium distachyon* and *O. sativa*, seemed to support this prediction [106], but more extensive taxonomic sampling now shows that gbM can be highly variable across species [4], even within the same taxonomic groups [3, 97]. The most extreme cases are the two angiosperm species that have no CMT3 (*E. salsugineum* and *C. planisiliqua*) and lack gbM altogether. Despite the loss of gbM, the transcriptional state of orthologous genes in these two species seems to be largely conserved, suggesting that gbM has no causal role in the functional conservation of these orthologs.

The emerging neutrality hypothesis

The potential uncoupling of gbM from transcriptional states has raised the question of why gbM often appears in moderately and highly expressed genes in the first place. An emerging hypothesis is that gbM is simply the neutral byproduct of active transcription. Bewick et al. [97] recently proposed a molecular model for this neutrality hypothesis in which H3K9me2 is stochastically incorporated into transcribed genes. The transient presence of H3K9me2 kickstarts CMT3-dependent methylation of CHG sites and occasionally leads to the methylation of neighboring CGs, which are then maintained by the CG methyltransferase MET1. However, not all transcribed genes are body methylated. Bewick et al. [97] identified additional sequence and chromatin features that facilitate the accumulation of gbM within transcribed genes.

The hypothesis that gbM is a neutral byproduct of transcription predicts that it should, at least partly, track the evolution of transcriptional states in plant populations, provided that the full DNA methylation machinery is in place. Preliminary evidence that supports this prediction comes from a recent integrative transcriptome and methylome analysis in *A. thaliana* natural accessions [108]. Causal modeling shows that most *cis*- or *trans*-acting SNPs that are associated with both the expression and the methylation levels of a given gene tend to affect methylation through their effects on gene expression rather than the other way around. In other words, methylation is a byproduct of genetic effects on transcription. Many of these causal SNPs show evidence of positive selection [73], suggesting that the evolving genetic basis that underlies these transcriptional states leaves secondary signatures at the level of gbM.

Methylome evolution over short timescales

As discussed above, our current state of knowledge points to genomic changes as a major cause of long-term methylome evolution. These genomic changes can be in the form of repeat or TE expansion or the result of genetic perturbations in pathways that control DNA methylation or transcriptional states. The species-level methylome footprints of these changes are expected to emerge gradually, as point or structural mutations need to arise first and then spread within or among populations through selection and drift (Fig. 3). An intriguing observation, however, is that heritable alterations in DNA methylation states can also emerge spontaneously and independently of genetic mutations [8, 57, 76–78, 109–113]. The most comprehensive demonstration of this phenomenon has come from the analysis of multi-generational methylome data from *A. thaliana* mutation accumulation lines (MA-lines) [76–78, 112]. Such lines are derived from a single founder plant (of the Columbia accession) and independently propagated for a large number of generations [114]. Detailed comparisons of the methylomes of MA-lines have been instrumental in providing the first estimates of the rate at which epimutations occur in plant genomes [76–78]. Efforts are now underway to try to understand the extent to which spontaneous epimutations contribute to methylome diversity in natural populations over short timescales up to thousands of generations.

Spontaneous epimutations can rapidly generate methylome diversity

Spontaneous epimutations can be defined as heritable stochastic changes in the methylation status of individual cytosines or of clusters of cytosines. In plants, such stochastic events can occur at CG, CHG, and CHH sites. The meiotic inheritance of epimutations, however, appears

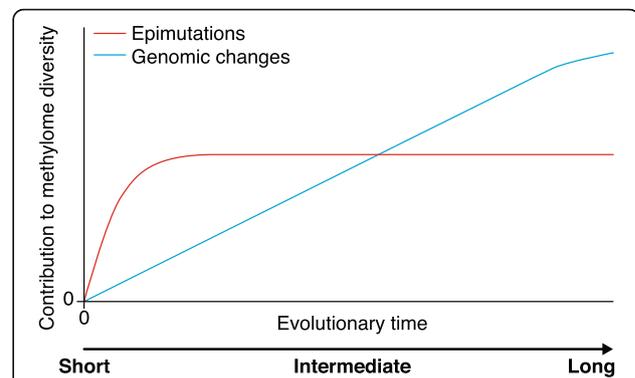


Fig. 3 We propose a heuristic model whereby genomic changes and spontaneous epimutations occur simultaneously, and contribute to interspecific or intraspecific methylome diversity over evolutionary time. For illustrative purposes, we assume that lineages descended from a common founder plant at time 0. The rapid accumulation of heritable epimutations dominates methylome diversity over short timescales but quickly converges to an equilibrium diversity value that is defined by the magnitude of forward and backward epimutation rates as well as by their ratios (i.e., the epimutation bias parameter). By contrast, genomic changes accumulate more gradually among lineages, and begin to dominate methylome diversity after longer evolutionary divergence times. An important empirical challenge is to delineate the relative contributions of these two processes based on methylome diversity data collected at any point along this evolutionary trajectory. Recent theoretical models for the analysis of the methylation site frequency spectrum (mSFS) provide an important step in this direction

to be mainly restricted to CG dinucleotides [76–78], probably as a result of context-specific methylation resetting during gametogenesis and early development [115]. Estimates in MA-lines indicate that the rate of forward epimutations (i.e., stochastic gains of methylation) is about 2.56×10^{-4} per CG site per haploid genome per generation, while the rate of backward epimutations (i.e., stochastic loss of methylation) is about 6.3×10^{-4} [78]. Hence, methylation loss is globally about 2.5 times as likely as methylation gain. The asymmetry in these rates has immediate consequences for understanding GMLs in *A. thaliana*: it implies that about 30% of all CG dinucleotides should be methylated at equilibrium and 70% unmethylated, provided that evolutionary forces such as selection and gene conversion are negligible. These percentages are roughly consistent with actual measurements of GMLs in the *A. thaliana* reference accession (Columbia), suggesting that epimutations are fundamental to methylome evolution despite the myriad of ways in which genomic changes can shape methylation patterns in the long term.

Putting these rates into perspective, the rate of CG epimutations is about five orders of magnitude higher than the rate of genetic mutations in *A. thaliana* (7×10^{-9}) [116]. In sheer numbers, about 10,000 CG methylation changes occur in a single generation, which contrasts with the two base changes resulting from genetic mutations.

The fast accumulation of these methylation changes causes methylomes to diverge rapidly over short time-scales. Even after only 30 generations of independent selfing, the methylomes of early-generation and late-generation MA-lines can be clearly distinguished. As the methylation status of individual CG sites is simultaneously subject to both forward and backward epimutations, methylome divergence does not increase linearly over time [72, 78, 117] but saturates rather quickly to some equilibrium divergence value (Fig. 3). On the basis of estimates from Van der Graaf et al. [78], only about 4000 generations would be needed in a hypothetical mutation accumulation experiment for methylome divergence to be within 99% of this value. This insight leads to the evolutionary prediction that epimutational processes should dominate methylome diversity in the early stages of lineage divergence but only marginally at later stages.

The high epimutation rates have additional implications for studying methylome diversity within and across populations. First, the observed shared methylated state between two individuals (so-called identity by state) cannot be assumed automatically to be inherited from the same parent (so-called identity by descent), because it could have been generated by independent epimutation events. This concept is defined as homoplasy and has been largely studied for microsatellite markers [118]. Second, as divergence in the methylome between populations increases rapidly, backward and forward epimutations would occur at many sites. Therefore, homoplasy will be observed when comparing diverged populations of the same species, thus decreasing the accuracy of inference of past evolutionary events.

Epimutation-induced methylome diversity patterns are potentially long-lived

Like genetic mutations, CG epimutations are not uniformly distributed across the genome, but vary in rate between different annotation contexts [76–78, 112]. In *A. thaliana*, the highest combined forward and backward rates are found in genes, with the forward rate (3.48×10^{-4}) being about four times lower than the backward rate (1.47×10^{-3}). In TEs, by contrast, these rates are much reduced, and the forward rate (3.24×10^{-4}) exceeds the backward rate (1.20×10^{-5}) by a factor of 30 [78]. The strong epimutation bias toward methylation gain in TEs is consistent with constitutive silencing of these sequences by DNA methylation. An important by-product of these annotation-specific rates (and their degree of asymmetry) is that some genomic regions diverge faster than others and also tend toward distinct equilibrium divergence values over time. That is, CG epimutations are expected to produce methylome diversity patterns along chromosomes that closely reflect the spatial distribution of various annotation units (i.e.,

chromosome architecture) (Fig. 4). In the *A. thaliana* MA-lines, this can be seen clearly when comparing pericentromeric regions (TE-rich) and chromosome arms (gene-rich), with the latter being on average about 2.3 times more divergent than the former (Fig. 4).

Because chromosome architecture is broadly stable over long evolutionary timescales, the signatures of epimutational events are potentially long-lived. Indeed, a striking observation is that the epimutation-induced methylome diversity patterns in the MA-lines are highly correlated with those seen among worldwide natural accessions (pericentromeric regions: $\rho = 0.94$, chromosome arms: $\rho = 0.72$; Fig. 4), despite the latter having diverged for hundreds of thousands of years [119, 120]. These correlations are even stronger, particularly in chromosome arms, when the MA-lines are compared to a selected sample of North American natural accessions that diverged from a common founder about 200 years ago [72] (pericentromeric regions: $\rho = 0.92$, chromosome arms: $\rho = 0.82$; Fig. 4). Together, these observations indicate that—while the accumulation of sequence polymorphisms affects methylation diversity patterns over time—in the current state of the species' evolutionary trajectory, these effects are not overwhelming. Similar conclusions can be reached on the basis of a careful evaluation of meQTL studies in *A. thaliana* accessions [63, 73, 75], which show that on average only about 18–35% of all DMRs are associated with *cis*- or *trans*-acting sequence polymorphisms [93]. The above insights raise the following important questions. Are spontaneous epimutations generally a major cause of methylome diversity in natural plant populations? And if so, what are the evolutionary forces that act on these epimutations?

Analysis of the methylation site frequency spectrum (mSFS)

One way to approach these questions is to analyze the mSFS (Fig. 5) using a theoretical model that explicitly accounts for forward and backward epimutations as well as for evolutionary forces such as selection and drift. Although this modeling approach goes back to Wright [121], results that are applicable for the analysis of genomic data have been obtained recently [122–124]. More popularized classic population genetics models that assume irreversible mutations (see also Wright [121]) on infinitely many sites [125], as is often the case for genomic data, are not suitable in the context of epimutations because of their reversibility and relatively high asymmetric rates. Recently, Charlesworth and Jain [123] derived analytical results based on the work of Wright [121], which incorporate many of the key features of epimutations (Box 1). Their formulas can be directly applied to WGBS-seq data that describe single methylation polymorphisms (SMPs) or DMRs to make inferences

Box 1 Analysis of the methylation site frequency spectrum (mSFS)

Consider a randomly mating, panmictic, diploid population with constant population size N . Each cytosine has two epiallelic states cM and cU , with the former denoting a methylated and the latter an unmethylated state. We assume that forward epimutations ($cU \rightarrow cM$) occur at rate $\alpha = 4N\mu_{UM}$, and backward epimutations ($cM \rightarrow cU$) at rate $\beta = 4N\mu_{MU}$. Selection acts with coefficient $\sigma = 2Ns$, where the relative fitness of the cU/cU and cM/cU epigenotypes over cM/cM are given by $1 + 2s$ and $1 + s$, respectively. According to Charlesworth and Jain [123] the probability that a sample of size n segregates for b cU variants (with $0 \leq b \leq n$) is

$$p_{n,b} = \binom{n}{b} \frac{F(\beta + b; \alpha + \beta + n; 2\sigma) \beta^{(b)} \alpha^{(n-b)}}{F(\beta, \alpha + \beta, 2\sigma) (\alpha + \beta)^{(n)}},$$

where $F(x,y,z)$ denotes the confluent hypergeometric function of the first kind and the d_{ij} are rising factorials [127]. Note that the equation has been slightly adapted to our notation. The proportion of segregating sites is $p_{seg} = 1 - p(0) - p(n)$ and the mSFS is obtained as

$$q_{n,b} = p_{n,b} / p_{seg}$$

We introduce this equation into a maximum likelihood framework to infer the epimutation rates and the selection coefficient from the observed mSFS, which can be constructed from whole genome bisulphite sequencing data. Assuming independent sites, the log-likelihood of a model M given data D is

$$\log [L(D; M)] = \sum_{b=1}^{n-1} d_{n,b} \log(q_{n,b}) + \text{constant},$$

Where $d_{n,b}$ is the observed number of sites at which the cU epiallele occurs b times in the sample, and $q_{n,b}$ is the probability that the cU epiallele occurs b times in the sample at a segregating site under model M [128]. To emphasize the proportion of the two epimutation rates α and β , we use the epimutation bias parameter r via $\beta = r\alpha$. Maximum likelihood estimates for the parameters r , α (thus β) and σ can be obtained by performing a grid search over the parameter space. The model with the highest likelihood is selected.

Note that the mSFS approach is also applicable when using ‘regions’ (i.e. clusters of cytosines) as units of analysis rather than individual cytosines. However, this shift in focus requires that differentially methylated regions (DMRs) can be assumed to exist in two epiallelic states (i.e. methylated and unmethylated) and that epimutation events occur at the level of ‘regions’.

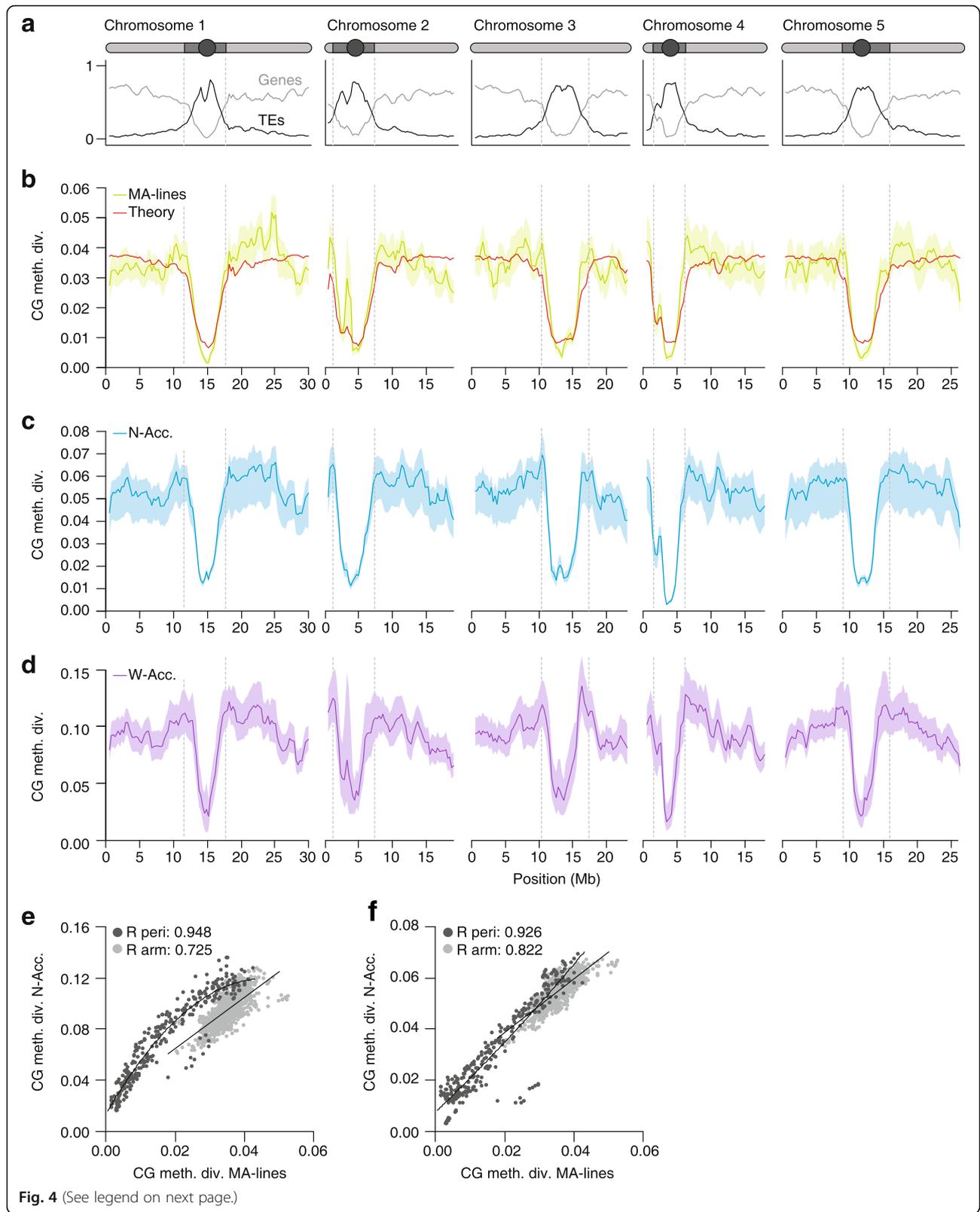
about the evolutionary role of epimutations and selection in shaping methylome diversity patterns in natural populations.

Analysis of mSFS in *A. thaliana*: an example

To demonstrate the power of this approach, we constructed the mSFS from public WGBS-seq data of 92 worldwide natural *A. thaliana* accessions [63] (Fig. 5; see Additional file 3 for a description of how the methylomes used for the mSFS calculations were filtered). These 92 accessions represent a so-called species-wide sample of *A. thaliana*, characterizing the collecting phase of the species’ coalescent tree [126]. This sample can be seen as a panmictic population and thus fulfills our model’s assumptions (Box 1). For this analysis, we focused only on genic CG sites, because this approach allowed us to draw connections between epimutational processes and the nature of gbM discussed above. As shown in Fig. 5, our theory-based estimates give an accurate description of the observed mSFS, indicating that the underlying model assumptions are sufficient and that epimutations are a major factor in shaping species-level methylome diversity in *A. thaliana*. Several important insights are emerging from this model fit.

First, the best fitting model provides no evidence for selection on genic CG epimutations at the genome-wide level. This observation is consistent with earlier theoretical models of the MA-lines, which have shown that epimutations accumulate neutrally under benign environmental conditions and in nearly isogenic sequence backgrounds [78]. The lack of selection also provides support to the molecular model of Bewick et al. [97], which posits that gbM is essentially a neutral by-product of expression, although a more detailed mSFS analysis that considers separate classes of BM and UM genes will be required to confirm this.

A second major insight from the mSFS fit is that the ratio of forward and backward population epimutation rates is similar to that estimated in the MA-lines (3.43 vs. 4.24, respectively). This result indicates that the epimutation bias parameter is robustly maintained in natural environments and in the context of varying genomic backgrounds, a conclusion that has also been reached by Hagmann et al. [72] using less formal arguments. Estimates of the actual epimutation rates, however, are not available from the mSFS output because the population epimutation parameters are a function of the effective population size (N_e), and cannot be disentangled (Box 1). This is unfortunate as it would be interesting to assess the extent to which the actual rates are modulated by environmental and genetic factors. A previous experiment in which MA-lines were derived under high-salinity soil conditions provided evidence



(See figure on previous page.)

Fig. 4 a Gene (*light gray*) and transposable element (*TE*) (*dark gray*) densities along the *A. thaliana* genome (Columbia reference). A schematic representation of the five chromosomes is shown above (*circle*, centromere; *dark gray*, pericentromeric region; *light gray*, arm). **b** Annotation-specific CG epimutations produce distinct methylome diversity (*CG meth. div.*) patterns among mutation accumulation lines (*MA-lines*) that have diverged for merely 30 generations (average diversity was calculated in 1 Mb sliding windows, step size 100 kb). These diversity patterns can be predicted from annotation-specific estimates of epimutation rate and the density distribution of annotation units along the genome (*red theoretical line*). **c** CG methylome diversity (*CG meth. div.*) patterns among 13 North American accessions (*N-Acc.*) (after around 200 generations of divergence). **d** Methylome diversity patterns among 138 worldwide accessions (*W-Acc.*) (after several hundred thousand years of divergence). **e** CG methylome diversity patterns are significantly correlated between the MA-lines and the W-Acc., both in pericentromeric (*peri*) regions (*dark gray dots*) as well as in euchromatic chromosome arms (*light gray dots*). **f** These correlations are even stronger when MA-lines are compared to the N-Acc., suggesting that the accumulation of DNA sequence polymorphism has perturbed epimutation-induced methylome diversity patterns over time

that epimutations are more frequent under this stressor [112]. Similar experiments are underway to assess the rate and spectrum of epimutations as a function of varying genomic backgrounds.

Interesting future directions in the analysis of mSFS

The mSFS analysis approach opens up exciting research avenues. Most notably, it provides a formal framework

for carrying out methylome-wide scans for signatures of epigenetic selection by identifying DMRs that significantly diverge from the expected mSFS. While the interpretation of such regions is difficult, as they could be the result of direct selection on methylation states or the outcome of indirect selection on *cis*- or *trans*-acting genetic variants, this approach would provide a way to prioritize regions for further analytical or experimental

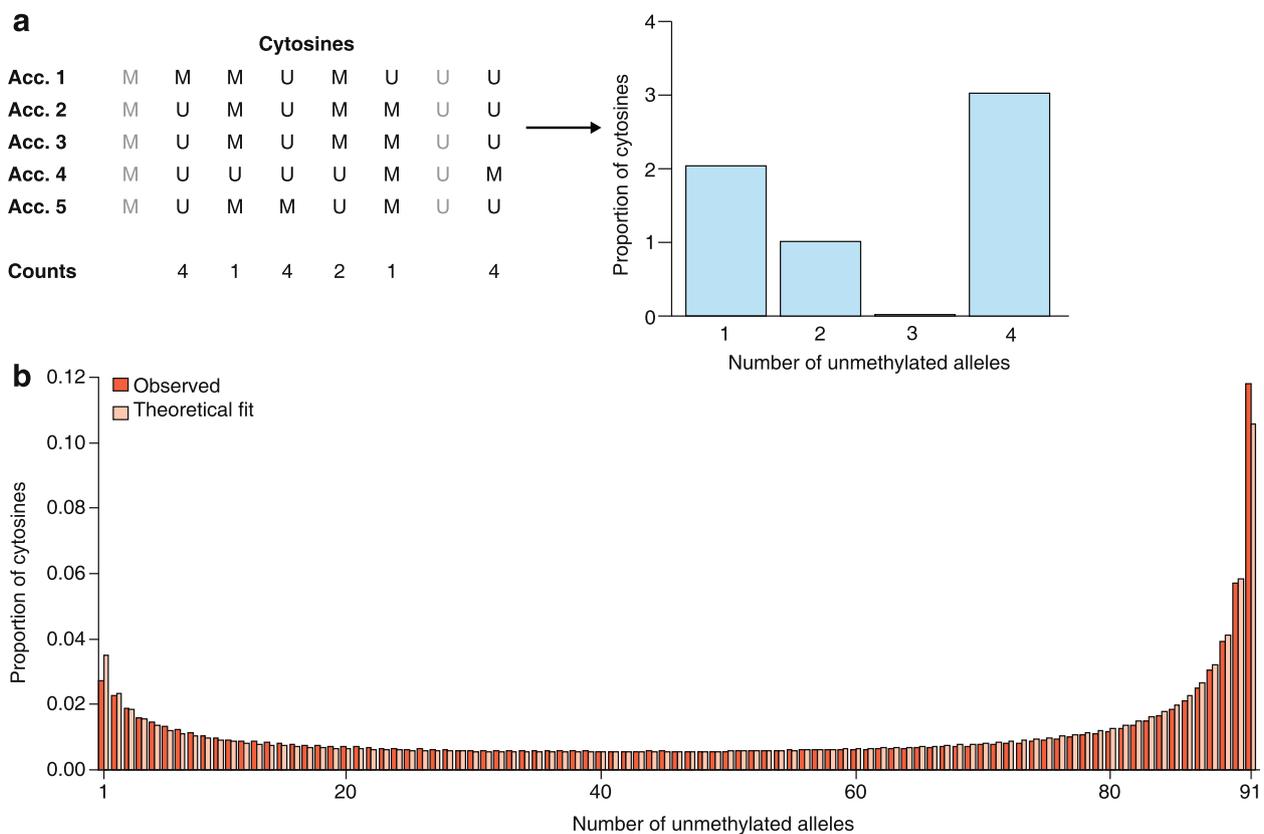


Fig. 5 a Simplification of the reconstruction of a methylation site frequency spectrum (mSFS). In this example, we consider a sample size of five accessions (*Acc.*), and eight sites among which two (in *gray*) are monomorphic and thus discarded for the mSFS. For each cytosine, each accession might exhibit a methylated (*M*) or an unmethylated (*U*) state. For the mSFS, counts are taken of the number of accessions that are unmethylated for that cytosine. These counts define discrete epiallelic classes (number of unmethylated alleles). **b** The observed frequencies of each epiallelic class is determined, in this case, from genic CG sites of 92 *A. thaliana* worldwide natural accessions (*red bars*), along with the maximum likelihood estimate based on the theoretical result of Charlesworth and Jain [123] (*pink bars*). The theoretical model (see Box 1) provides an accurate fit to the observed genic CG methylation diversity patterns, suggesting that CG epimutations are a major factor in shaping methylome diversity in natural populations of *A. thaliana* over evolutionary timescales

analysis. These methylome-wide scans will also provide a new perspective on the large number of methylomes that have been recently collected in *A. thaliana*, or will be collected for other plant species in the near future. Another interesting extension of the mSFS approach is to generalize the theoretical result of Charlesworth and Jain [123] to account for time dependence and therefore to incorporate changes in the population size. Such a model could be used in conjunction with genic CG mSFS data to define a kind of ‘fast-ticking’ molecular clock. Genic CG epimutations can be considered as neutral and occur at rates far exceeding the genetic mutation rate, and so such a re-calibrated clock would yield high-resolution insights into very recent demographic events that would otherwise be invisible on the basis of DNA sequencing data alone.

Conclusions

The recent availability of high-resolution inter- and intraspecific methylome data is providing new insights into the evolutionary role of DNA methylation in plants. Such insights complement the tremendous progress made in recent years in understanding more proximal questions regarding the molecular mechanisms that control DNA methylation during the life course of a plant and during its reproductive stages. This review provides a first unified framework for understanding the evolution of methylation in plants, based on the fact that the epigenomic divergence observed at the longer timescales is necessarily the result of processes occurring within populations at shorter timescales.

At the population level, spontaneous epimutations appear to be a major factor in generating methylome diversity. These epimutations are characterized by their high, asymmetric rates, and the fact that they occur at a finite number of cytosines. Following population genetics theory, drift and selection should drive the changes in epimutation frequencies over time, thus shaping the mSFS in a population. We predict that most plant populations will be close to statistical equilibrium with respect to epimutation, genetic drift, and selection, and that they will be characterized by extensive homoplasy. Cases of positive or purifying selection on epialleles have never been reported, probably because of a lack of appropriate statistical analyses. Hence, an open question is whether epigenetic selection is pervasive or rare in plant populations. A theory-based analysis of the empirical mSFS provides a framework for detecting signatures of positive and purifying selection at the genome-wide scale. Using such an approach, future studies should assess the extent to which the mSFS for different annotation units is conserved between plant species. For instance, is the neutral mSFS that we have detected in *A. thaliana* natural populations typical? The fact that genic

sequences in complex genomes are often ‘contaminated’ with TEs and sequence repeats [4] would suggest that epimutation dynamics differ fundamentally among different genomes and may be subject to selection. Population-level methylome data in several other plant species will soon emerge to answer these questions.

When populations diverge, drift and high epimutation rates generate fast divergence in methylation at existing cytosine sites. If local adaptation occurs and is mediated by DNA methylation, selection should be observable in the mSFS, and possibly also with the greater divergence between populations of mSFS in key genes for adaptation. Within populations, more drastic genomic changes will arise slowly; these might include, for example, genome rearrangements, gene duplication, the repeating or expansion of TEs, changes in methylation pathways, and so on. We know that these genomic changes affect methylation patterns because DMRs are often associated with segregating structural variants or with mutations in methyltransferase genes. When these features become fixed in a population, the methylome landscape changes drastically. This can be then observed in comparative epigenomics studies that show the cumulative outcome of genetic changes.

From a theoretical perspective, a crucial future step is to develop models that bridge these different time and spatial scales. Such models should include not only population genetic processes (drift, epimutation, recombination, migration, and selection) but also genomic rearrangements and TE dynamics to derive testable hypotheses and statistics suited for the analysis of intra- and interpopulation and species data.

These data-driven modeling efforts should be complemented by rigorous experimental studies that determine how heritable DNA methylation changes arise in different plant species and mating systems, and the extent to which these changes contribute to plant fitness and respond to artificial or natural selection.

Additional files

Additional file 1: Plant species whose methylomes have been analyzed by whole-genome bisulfite sequencing (WGBS-seq) or by high-performance liquid chromatography (HPLC). (PDF 277 kb)

Additional file 2: Figure S1. GMLs of different taxa measured by HPLC and WGBS-seq. **Figure S2.** Correlation between genome size and total number of repeats in the genome. (PDF 1044 kb)

Additional file 3: Filtering of the methylomes used for the calculation of the mSFS. (DOCX 17 kb)

Abbreviations

BM: Body methylated; CMT3: CHROMOMETHYLASE 3; DMR: Differentially methylated region; gbM: Gene-body methylation; GML: Genome-wide methylation level; H3K9me2: Histone H3 lysine 9 di-methylation; HPLC: High-performance liquid chromatography; K_A : Number of non-synonymous substitutions per non-synonymous sites; K_{INT} : Intron divergence; K_S : Number of

synonymous substitutions per synonymous sites; LD: Linkage disequilibrium; MA-line: Mutation accumulation line; meQTL: Methylation quantitative trait loci; mSFS: Methylation site frequency spectrum; SNP: Single nucleotide polymorphism; TE: Transposable element; UM: Unmethylated; WGBS-seq: Whole-genome bisulfite sequencing

Acknowledgements

We thank R.J. Schmitz, C.E. Niederhuth, and C. Alonso for their help in re-analyzing their published data.

Funding

FJ and DR acknowledge support from the Technical University of Munich-Institute for Advanced Study funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement #291763. AT and DZ acknowledge funding from the Deutsche Forschungsgemeinschaft grants TE 809/6-2 and STE 325/14.

Authors' contributions

AV, DZ, RW, and DR analyzed the data. DZ and AT developed the statistical analysis of the model. FJ wrote the paper with input from all authors. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Population Epigenetics and Epigenomics, Technical University of Munich, Liesel-Beckman-Str. 2, 85354 Freising, Germany. ²Population Genetics, Technical University of Munich, Liesel-Beckman-Str. 2, 85354 Freising, Germany. ³Groningen Bioinformatics Centre, University of Groningen, 9747 AG Groningen, The Netherlands. ⁴Institute for Advanced Study, Technical University of Munich, Lichtenbergstr. 2a, 85748 Garching, Germany.

Published online: 20 December 2016

References

- Feng S. Epigenetic reprogramming in plant and animal development. *Science*. 2010;330:622–7.
- Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*. 2010;328:916–9.
- Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, et al. Widespread natural variation of DNA methylation within angiosperms. *Genome Biol*. 2016;17:194.
- Takuno S, Ran J-H, Gaut BS. Evolutionary patterns of genic DNA methylation vary across land plants. *Nat Plants*. 2016;2:15222.
- Das OP, Messing J. Variegated phenotype and developmental methylation changes of a maize allele originating from epimutation. *Genetics*. 1994;136:1121–41.
- Bender J, Fink GR. Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*. *Cell*. 1995;83:725–34.
- Melquist S, Bender J. Transcription from an upstream promoter controls methylation signaling from an inverted repeat of endogenous genes in *Arabidopsis*. *Genes Dev*. 2003;17:2036–47.
- Ong-Abdullah M, Ordway JM, Jiang N, Ooi S-E, Kok S, Sarpan N, et al. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature*. 2015;525:533–7.
- Banks JA, Masson P, Fedoroff N. Molecular mechanisms in the developmental regulation of the maize Suppressor-Mutator transposable element. *Genes Dev*. 1988;2:1364–80.
- Jacobsen SE. Hypermethylated SUPERMAN epigenetic alleles in *Arabidopsis*. *Science*. 1997;277:1100–3.
- Soppe WJJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, et al. The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell*. 2000;6:791–802.
- Stam M, Belele C, Dorweiler JE, Chandler VL. Differential chromatin structure within a tandem array 100 kb upstream of the maize *b1* locus is associated with paramutation. *Genes Dev*. 2002;16:1906–18.
- Colot V, Maloisel L, Rossignol JL. Interchromosomal transfer of epigenetic states in *Ascombolus*: transfer of DNA methylation is mechanistically related to homologous recombination. *Cell*. 1996;86:855–64.
- Stokes TL, Kunkel BN, Richards EJ. Epigenetic variation in *Arabidopsis* disease resistance. *Genes Dev*. 2002;16:171–82.
- Quadrana L, Almeida J, Asis R, Duffy T, Dominguez PG, Bermúdez L, et al. Natural occurring epialleles determine vitamin E accumulation in tomato fruits. *Nat Commun*. 2014;5:3027.
- Reinders J, Wulff BBH, Mirouze M, Mari-Ordóñez A, Dapp M, Rozhon W, et al. Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev*. 2009;23:939–50.
- Johannes F, Porcher E, Teixeira FK, Saliba-Colombani V, Simon M, Agier N, et al. Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genet*. 2009;5:e1000530.
- Roux F, Colomé-Tatché M, Edelist C, Wardenaar R, Guerche P, Hospital F, et al. Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. *Genetics*. 2011;188:1015–7.
- Eichten SR, Schmitz RJ, Springer NM. Epigenetics: beyond chromatin modifications and complex genetic regulation. *Plant Physiol*. 2014;165:933–47.
- Silveira AB, Trontin C, Cortijo S, Barau J, Del Bem LEV, Loudet O, et al. Extensive natural epigenetic variation at a de novo originated gene. *PLoS Genet*. 2013;9:3–10.
- Tsukahara S, Kobayashi A, Kawabe A, Mathieu O, Miura A, Kakutani T. Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature*. 2009;461:423–6.
- Mirouze M, Reinders J, Bucher E, Nishimura T, Schneeberger K, Ossowski S, et al. Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature*. 2009;461:1–5.
- Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature*. 2001;411:212–4.
- Singer T, Jordan C, Martienssen RA. Robertson's *Mutator* transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA Methylation (DDM1)*. *Genes Dev*. 2001;15:591–602.
- Cheng C, Tarutani Y, Miyao A, Ito T, Yamazaki M, Sakai H, et al. Loss of function mutations in the rice chromomethylase OsCMT3a cause a burst of transposition. *Plant J*. 2015;83:1069–81.
- Kim KD, El Baidouri M, Abernathy B, Iwata-Otsubo A, Chavarro C, Gonzales M, et al. A comparative epigenomic analysis of polyploidy-derived genes in soybean and common bean. *Plant Physiol*. 2015;168:1433–47.
- Maloisel L, Rossignol JL. Suppression of crossing-over by DNA methylation in *Ascombolus*. *Genes Dev*. 1998;12:1381–9.
- Mirouze M, Lieberman-Lazarovich M, Aversano R, Bucher E, Nicolet J, Reinders J, et al. Loss of DNA methylation affects the recombination landscape in *Arabidopsis*. *Proc Natl Acad Sci U S A*. 2012;109:5880–5.
- Colomé-Tatché M, Cortijo S, Wardenaar R, Morgado L, Lahouze B, Sarazin A, et al. Features of the *Arabidopsis* recombination landscape resulting from the combined loss of sequence variation and DNA methylation. *Proc Natl Acad Sci U S A*. 2012;109:16240–5.
- Melamed-Bessudo C, Levy AA. Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in *Arabidopsis*. *Proc Natl Acad Sci U S A*. 2012;109:E981–8.
- Yelina NE, Choi K, Chelysheva L, Macaulay M, de Snoo B, Wijnker E, et al. Epigenetic remodeling of meiotic crossover frequency in *Arabidopsis thaliana* DNA methyltransferase mutants. *PLoS Genet*. 2012;8:e1002844.
- Shen H, He H, Li J, Chen W, Wang X, Guo L, et al. Genome-wide analysis of DNA methylation and gene expression changes in two *Arabidopsis* ecotypes and their reciprocal hybrids. *Plant Cell*. 2012;24:875–92.
- Dapp M, Reinders J, Bédiée A, Balsera C, Bucher E, Theiler G, et al. Heterosis and inbreeding depression of epigenetic *Arabidopsis* hybrids. *Nat Plants*. 2015;1:15092.
- Rigal M, Becker C, Pélissier T, Pogorelnik R, Devos J, Ikeda Y, et al. Epigenome confrontation triggers immediate reprogramming of DNA methylation and transposon silencing in *Arabidopsis thaliana* F1 epihybrids. *Proc Natl Acad Sci U S A*. 2016;113:E2083–92.
- Lauss K, Wardenaar R, van Hulten MHA, Guryev V, Keurentjes JJB, Stam M, et al. Epigenetic divergence is sufficient to trigger heterosis in *Arabidopsis thaliana*. *bioRxiv*. 2016; doi:http://dx.doi.org/10.1101/059980.
- Groszmann M, Greaves IK, Fujimoto R, Peacock WJ, Dennis ES. The role of epigenetics in hybrid vigour. *Trends Genet*. 2013;29:684–90.
- Chen ZJ. Genomic and epigenetic insights into the molecular bases of heterosis. *Nat Rev Genet*. 2013;14:471–82.
- Kirkbride RC, Yu HH, Nah G, Zhang C, Shi X, Chen ZJ. An epigenetic role for disrupted paternal gene expression in postzygotic seed abortion in *Arabidopsis* interspecific hybrids. *Mol Plant*. 2015;8:1766–75.

39. Fort A, Ryder P, Mckeown PC, Wijnen C, Aarts MG, Sulpice R, et al. Disaggregating polyploidy, parental genome dosage and hybridity contributions to heterosis in *Arabidopsis thaliana*. *New Phytol.* 2016;209:590–9.
40. Groszmann M, Gonzalez-Bayon R, Lyons RL, Greaves IK, Kazan K, Peacock WJ, et al. Hormone-regulated defense and stress response networks contribute to heterosis in *Arabidopsis* F1 hybrids. *Proc Natl Acad Sci U S A.* 2015;112: E6397–406.
41. Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, et al. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *Elife.* 2015;4:e09343.
42. Feil R, Fraga MF. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet.* 2012;13:97–109.
43. Meyer P. Epigenetic variation and environmental change. *J Exp Bot.* 2015;66:3541–8.
44. Zhang X. Dynamic differential methylation facilitates pathogen stress response in *Arabidopsis*. *Proc Natl Acad Sci U S A.* 2012;109:12842–3.
45. Yu A, Lepere G, Jay F, Wang J, Bapaume L, Wang Y, et al. Dynamics and biological relevance of DNA demethylation in *Arabidopsis* antibacterial defense. *Proc Natl Acad Sci U S A.* 2013;110:2389–94.
46. López Sánchez A, Stassen JH, Furci L, Smith LM, Ton J. The role of DNA (de)methylation in immune responsiveness of *Arabidopsis*. *Plant J.* 2016; 88:361–74.
47. Espinas NA, Saze H, Saijo Y. Epigenetic control of defense signaling and priming in plants. *Front Plant Sci.* 2016;7:1201.
48. Luna E, Ton J. The epigenetic machinery controlling transgenerational systemic acquired resistance. *Plant Signal Behav.* 2012;7:615–8.
49. Conrath U, Beckers GJM, Langenbach CJG, Jaskiewicz MR. Priming for enhanced defense. *Annu Rev Phytopathol.* 2015;53:97–119.
50. Rapp RA, Wendel JF. Epigenetics and plant evolution. *New Phytol.* 2005; 168:81–91.
51. Richards EJ, Reinders J, Wulff BBH, Mirouze M. Quantitative epigenetics: DNA sequence variation need not apply. *Genes Dev.* 2009;23:1601–5.
52. Weigel D, Colot V. Epialleles in plant evolution. *Genome Biol.* 2012;13:249.
53. Diez CM, Roessler K, Gaut BS. Epigenetics and plant genome evolution. *Curr Opin Plant Biol.* 2014;18:1–8.
54. Springer NM. Epigenetics and crop improvement. *Trends Genet.* 2013;29:241–7.
55. Ji L, Neumann DA, Schmitz RJ. Crop epigenomics: identifying, unlocking, and harnessing cryptic variation in crop genomes. *Mol Plant.* 2014;8:860–70.
56. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet.* 2010;11:204–20.
57. Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell.* 2013;152:352–64.
58. Laird PW. Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet.* 2010;11:191–203.
59. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell.* 2006;126:1189–201.
60. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature.* 2008;452:215–9.
61. Lister R, Malley RCO, Tonti-filippini J, Gregory BD, Berry CC, Miller AH, et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell.* 2008;133:523–36.
62. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet.* 2007;39:61–9.
63. Schmitz RJ, Schultz MD, Urlich MA, Nery JR, Pelizzola M, Libiger O, et al. Patterns of population epigenomic diversity. *Nature.* 2013;495:193–8.
64. Gent JJ, Ellis NA, Guo L, Harkess AE, Yao Y, Zhang X, et al. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res.* 2013;23:628–37.
65. Seymour DK, Koenig D, Hagemann J, Becker C, Weigel D. Evolution of DNA methylation patterns in the *Brassicaceae* is driven by differences in genome organization. *PLoS Genet.* 2014;10:e1004785.
66. Amborella Genome Project, Albert VA, Barbazuk WB, dePamphilis CW, Der JP, Leebens-Mack J, et al. The *Amborella* genome and the evolution of flowering plants. *Science.* 2013;342:1241089.
67. Zhong S, Fei Z, Chen Y-R, Zheng Y, Huang M, Vrebalov J, et al. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. *Nat Biotechnol.* 2013;31:154–9.
68. Alonso C, Pérez R, Bazaga P, Herrera CM. Global DNA cytosine methylation as an evolving trait: phylogenetic signal and correlated evolution with genome size in angiosperms. *Front Genet.* 2015;5:1–9.
69. Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, et al. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet.* 2011;43:956–63.
70. Gan X, Stegle O, Behr J, Steffen JG, Drewe P, Hildebrand KL, et al. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. *Nature.* 2011;477:419–23.
71. Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, Platzer A, et al. Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat Genet.* 2013;45:884–90.
72. Hagemann J, Becker C, Müller J, Stegle O, Meyer RC, Wang G, et al. Century-scale methylome stability in a recently diverged *Arabidopsis thaliana* lineage. *PLoS Genet.* 2015;11:e1004920.
73. Dubin MJ, Zhang P, Meng D, Remigereau M-S, Osborne EJ, Paolo Casale F, et al. DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *Elife.* 2015;4:e05255.
74. 1001 Genomes Consortium. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell.* 2016;166:481–91.
75. Kawakatsu T, Huang SC, Jupe F, Sasaki E, Schmitz RJ, Urlich MA, et al. Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell.* 2016;166:492–505.
76. Becker C, Hagemann J, Müller J, Koenig D, Stegle O, Borgwardt K, et al. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature.* 2011;480:245–9.
77. Schmitz RJ, Schultz MD, Lewsey MG, Malley RCO, Urlich MA, Libiger O, et al. Transgenerational epigenetic instability is a source of novel methylation variants. *Science.* 2011;334:369–73.
78. Van der Graaf A, Wardenaar R, Neumann DA, Taudt A, Shaw RG, Jansen RC, et al. Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proc Natl Acad Sci U S A.* 2015;112:6676–81.
79. Schmitz RJ, He Y, Valdés-López O, Res G, Gent JJ, Ellis NA, et al. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. *Genome Res.* 2013;23:1663–74.
80. Virdi KS, Laurie JD, Xu Y-Z, Yu J, Shao M-R, Sanchez R, et al. *Arabidopsis* MSH1 mutation alters the epigenome and produces heritable changes in plant growth. *Nat Commun.* 2015;6:6386.
81. West PT, Li Q, Ji L, Eichten SR, Song J, Vaughn MW, et al. Genomic distribution of H3K9me2 and DNA methylation in a maize genome. *PLoS One.* 2014;9:1–10.
82. Li Q, Eichten SR, Hermanson PJ, Zaunbrecher VM, Song J, Wendt J, et al. Genetic perturbation of the maize methylome. *Plant Cell.* 2014;26:4602–16.
83. Li X, Zhu C, Yeh CT, Wu W, Takacs EM, Petsch KA, et al. Genic and nongenic contributions to natural variation of quantitative traits in maize. *Genome Res.* 2012;22:2436–44.
84. Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol.* 2014;65:505–30.
85. Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. *Genome Biol.* 2016;17:37.
86. Bennetzen JL, Ma J, Devos KM. Mechanisms of recent genome size variation in flowering plants. *Ann Bot.* 2005;95:127–32.
87. Woo HR, Richards EJ. Natural variation in DNA methylation in ribosomal RNA genes of *Arabidopsis thaliana*. *BMC Plant Biol.* 2008;8:92.
88. Quadrona L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddloh JA, et al. The *Arabidopsis thaliana* mobilome and its impact at the species level. *Elife.* 2016;5:e15716.
89. Stuart T, Eichten SR, Cahn J, Karpievitch Y, Borevitz JO, Lister R. Population scale mapping of novel transposable element diversity reveals links to gene regulation and epigenomic variation. *bioRxiv.* 2016; doi: <http://dx.doi.org/10.1101/039511>.
90. Mari-Ordóñez A, Marchais A, Etcheverry M, Martin A, Colot V, Voinnet O. Reconstructing de novo silencing of an active plant retrotransposon. *Nat Genet.* 2013;45:1029–39.
91. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* 2009;19:1419–28.
92. Ahmed I, Sarazin A, Bowler C, Colot V, Quesneville H. Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in *Arabidopsis*. *Nucleic Acids Res.* 2011;39:6919–31.

93. Tautd A, Colomé-Tatché M, Johannes F. Genetic sources of population epigenomic variation. *Nat Rev Genet.* 2016;17:319–32.
94. Eichten SR, Briskine R, Song J, Li Q, Swanson-Wagner R, Hermanson PJ, et al. Epigenetic and genetic influences on DNA methylation variation in maize populations. *Plant Cell.* 2013;25:2783–97.
95. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, et al. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat Struct Mol Biol.* 2014;21:64–72.
96. Hu L, Li N, Xu C, Zhong S, Lin X, Yang J, et al. Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proc Natl Acad Sci U S A.* 2014;111:10642–7.
97. Bewick AJ, Ji L, Niederhuth CE, Willing E-M, Hofmeister BT, Shi X, et al. On the origin and evolutionary consequences of gene body DNA methylation. *Proc Natl Acad Sci U S A.* 2016;113:9111–6.
98. Lindroth AM, Cao X, Jackson JP, Zilberman D, McCallum CM, Henikoff S, et al. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science.* 2001;292:2077–80.
99. Shen X, De Jonge J, Forsberg SKG, Pettersson ME, Sheng Z, Hennig L, et al. Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. *PLoS Genet.* 2014;10:e1004842.
100. Jullien PE, Susaki D, Yelagandula R, Higashiyama T, Berger F. DNA methylation dynamics during sexual reproduction in *Arabidopsis thaliana*. *Curr Biol.* 2012;22:1825–30.
101. Papa CM, Springer NM, Muszynski MG, Meeley R, Kaepler SM. Maize chromomethylase *Zea methyltransferase2* is required for CpNpG methylation. *Plant Cell.* 2001;13:1919–28.
102. Matzke MA, Kanno T, Matzke AJM. RNA-directed DNA methylation: the evolution of a complex epigenetic pathway in flowering plants. *Annu Rev Plant Biol.* 2014;66:1–25.
103. Bewick AJ, Niederhuth CE, Rohr NA, Griffin PT, Leebens-Mack J, Schmitz RJ. The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants. *bioRxiv.* 2016; doi: <http://dx.doi.org/10.1101/054924>.
104. Willing E-M, Rawat V, Mandáková T, Maumus F, James GV, Nordström KJV, et al. Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. *Nat Plants.* 2015;1:14023.
105. Takuno S, Gaut BS. Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly. *Mol Biol Evol.* 2012;29:219–27.
106. Takuno S, Gaut BS. Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proc Natl Acad Sci U S A.* 2013;110:1797–802.
107. Genereux DP, Miner BE, Bergstrom CT, Laird CD. A population-epigenetic model to infer site-specific methylation rates from double-stranded DNA methylation patterns. *Proc Natl Acad Sci U S A.* 2005;102:5802–7.
108. Meng D, Dubin M, Zhang P, Osborne EJ, Stegle O, Clark RM, et al. Limited contribution of DNA methylation variation to expression regulation in *Arabidopsis thaliana*. *PLoS Genet.* 2016;12:e1006141.
109. Cubas P, Vincent C, Coen E. An epigenetic mutation responsible for natural variation in floral symmetry. *Nature.* 1999;401:157–61.
110. Manning K, Tör M, Poole M, Hong Y, Thompson AJ, King GJ, et al. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet.* 2006;38:948–52.
111. Eichten SR, Swanson-Wagner RA, Schnable JC, Waters AJ, Hermanson PJ, Liu S, et al. Heritable epigenetic variation among maize inbreds. *PLoS Genet.* 2011;7:e1002372.
112. Jiang C, Mithani A, Belfield EJ, Mott R, Hurst LD, Harberd NP. Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. *Genome Res.* 2014;24:1821–9.
113. Verhoeven KJF, Jansen JJ, van Dijk PJ, Biere A. Stress-induced DNA methylation changes and their heritability in asexual dandelions. *New Phytol.* 2010;185:1108–18.
114. Shaw RG, Byers DL, Darms E. Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. *Genetics.* 2000;155:369–78.
115. Kawashima T, Berger F. Epigenetic reprogramming in plant sexual reproduction. *Nat Rev Genet.* 2014;15:613–24.
116. Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, et al. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science.* 2010;327:92–4.
117. Becker C, Weigel D. Epigenetic variation: origin and transgenerational inheritance. *Curr Opin Plant Biol.* 2012;15:562–7.
118. Estoup A, Jarne P, Cornuet JM. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol.* 2002;11:1591–604.
119. Novikova PY, Hohmann N, Nizhynska V, Tsuchimatsu T, Ali J, Muir G, et al. Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nat Genet.* 2016;48:1077–82.
120. Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, et al. The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol.* 2005;3:1289–99.
121. Wright S. Evolution in mendelian populations. *Genetics.* 1931;16:97–159.
122. Song YS, Steinrücken M. A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics.* 2012;190:1117–29.
123. Charlesworth B, Jain K. Purifying selection, drift, and reversible mutation with arbitrarily high mutation rates. *Genetics.* 2014;198:1587–602.
124. Wang J, Fan C. A neutrality test for detecting selection on DNA methylation using single methylation polymorphism frequency spectrum. *Genome Biol Evol.* 2014;7:154–71.
125. Kimura M. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics.* 1969;61:893–903.
126. Wakeley J, Aliacar N. Gene genealogies in a metapopulation. *Genetics.* 2001; 159:893–905.
127. Abramowitz M, Stegun IA. Handbook of mathematical functions. New York: Dover; 1965.
128. Živković D, Steinrücken M, Song YS, Stephan W. Transition densities and sample frequency spectra of diffusion processes with selection and variable population size. *Genetics.* 2015;200:601–17.