Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

Stochasticity in gene body methylation

Constantin Goeldel and Frank Johannes

Abstract

Gene body methylation (gbM) is a widely conserved epigenetic feature of plant genomes. Efforts to delineate the mechanisms by which gbM contributes to transcriptional regulation remain largely inconclusive, and its evolutionary significance continues to be debated. Curiously, although steady-state gbM levels are remarkably stable across mitotic and meiotic cell divisions, the methylation status of individual CG dinucleotides in gbM genes is highly stochastic. How can these two seemingly contradictory observations be reconciled? Here, we discuss how stochastic processes relate to gbM maintenance dynamics. We show that a quantitative understanding of these processes can shed deeper insights into the molecular and evolutionary biology of this enigmatic epigenetic trait.

Addresses

Plant Epigenomics, Technical University of Munich, Germany

Corresponding author: Johannes, Frank (f.johannes@tum.de)

Current Opinion in Plant Biology 2023, **75:**102436

This review comes from a themed issue on **Epigenetics and gene regulation**

Edited by **Nathan Springer** and **Hidetoshi Saze**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.pbi.2023.102436

Keywords

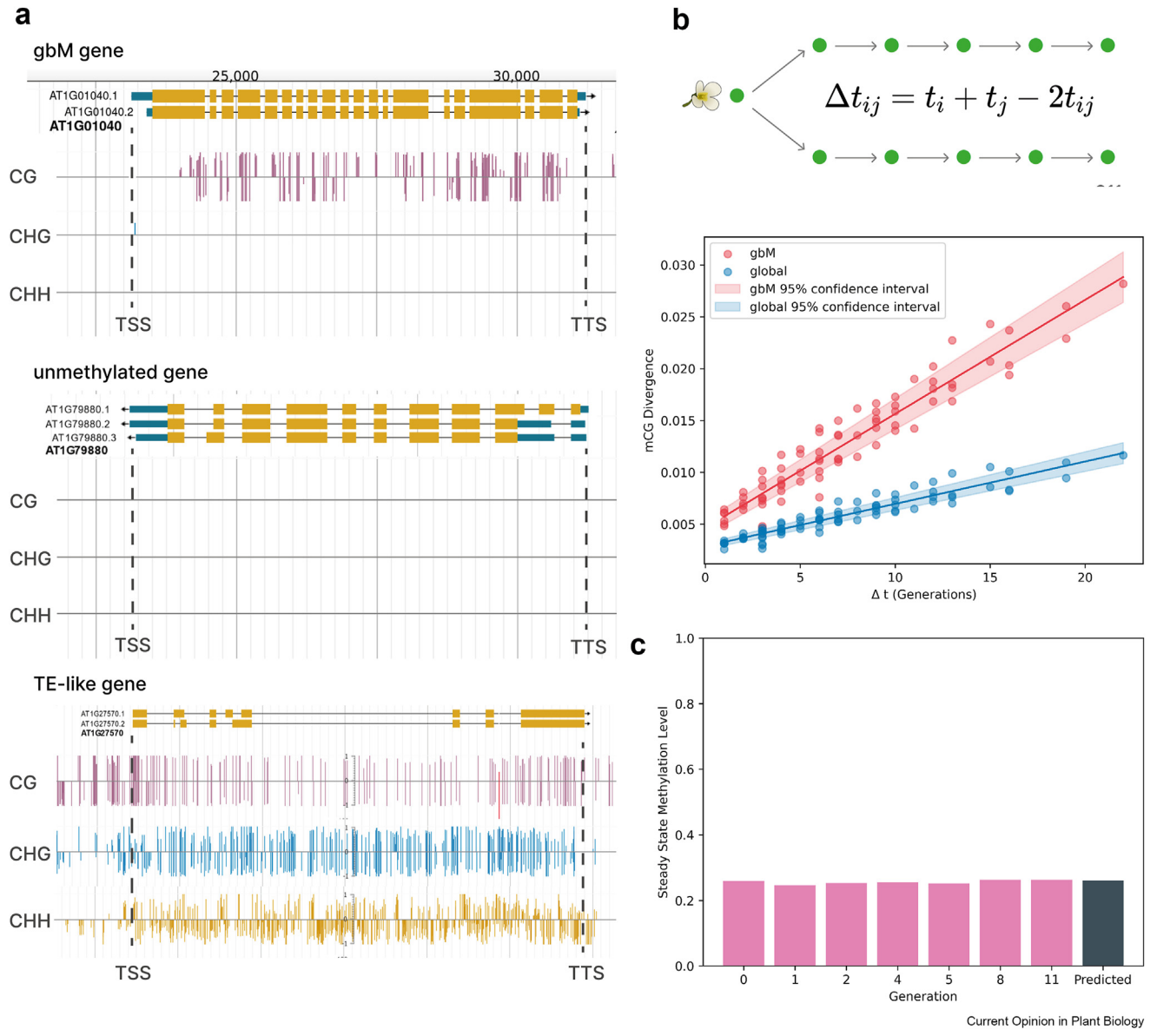
DNA methylation, Gene body methylation, Epimutations, Epimutation rate, MET1, H2A.Z, CMT3, Steady state, Modeling, Evolution.

The genomes of most angiosperms, but also those of some ferns and gymnosperms, feature a subclass of genes that are extensively methylated on CG dinucleotides, but not on cytosines in sequence contexts CHG and CHH (where H = A, T, C) [1–5]. These so-called gene body methylated (gbM) genes differ from other genes, where cytosine methylation is either completely absent or present in all three sequence contexts simultaneously [6] (Figure 1a). GbM genes are enriched in housekeeping functions, are moderately expressed, and display low transcriptional variability across cells and tissues [7–10] (extensively reviewed in Refs. [6,11]). Because of their essential cellular functions, gbM genes are evolutionarily conserved at the nucleotide level [2,3,12–15] and tend to evolve slowly [16]. Interestingly, the conservation of gbM levels is similarly strong across various plant species [2,3,17–19]. This has led to the hypothesis that gbM plays a crucial role in transcriptional regulation [8,20,21] and is potentially under selection [22,23]. However, efforts to identify specific regulatory mechanisms have yielded inconsistent results, and their functional significance is a matter of debate [6,11]. Here, we highlight that quantitative insights into the maintenance dynamics of gbM in plants can contribute to a deeper molecular and evolutionary understanding of this enigmatic epigenetic trait.

gbM is simultaneously stable and highly dynamic

In line with their evolutionary preservation, gbM levels (measured as an average across genes) remain remarkably stable over mitotic and meiotic cell divisions [24–27] (Figure 1c). However, this is not the case when individual CG sites within gbM genes are analyzed [24,28,29]. Observations over multiple generations indicate that stochastic methylation changes at individual CGs, known as "spontaneous epimutations" [30–32], occur at a remarkably high rate ($\sim 10^{-3}$ to 10^{-4} per haploid CG per generation) [24,25,28,30,33], and accumulate in gbM genes much more rapidly than in the genome on average (Figure 1b). Although these insights seem contradictory, they do suggest that—while the average gbM level is tightly regulated—the methylation status of individual CGs can "fluctuate" as long as the proportion of methylated to unmethylated cytosines remains constant over time. The relationship between steady-state gbM levels and the rates of stochastic methylation gain (α) and loss (β) at individual CGs can be well described mathematically [34,35] (Box 1). Statistical estimates of these rates are sufficient to accurately predict steady-state gbM levels without any auxiliary information (Figure 1c). Interestingly, recent work further shows that the stochastic process of gaining or losing CG methylation within gbM genes is highly robust [25]. That is, with some notable exceptions (below), the gain and loss rates, themselves, are not easily perturbed by different genotypes or multigenerational environmental stressors [25]. Whether this robustness is the cause or the consequence of steady-state stability of gbM levels remains to be determined.

Figure 1



DNA methylation patterns and epimutation accumulation in gbM genes. (a) DNA methylation patterns in gbM genes, unmethylated genes (UM) and transposable element (TE)-like genes. Screenshot from A. thaliana epigenome browser [62]. GbM genes are enriched in CG methylation and depleted in CHG and CHH methylation; UM genes carry no methylation in any context; TE-like genes are methylated in all three contexts. **(b)** Pedigree of the selfing-derived mutation accumulation (MA) lines from Shahryary et al. [35]. Only generations (G0 to GN) from which samples were analyzed with WGBS are shown. The divergence time between two samples is given by the summed generational distance of each sample to the most recent common ancestor. Lower: Dependence of mCG divergence on divergence time between any two samples. Divergence increases according to a neutral epimutation process and depends only on the stochastic methylation gain (α) and loss (β) rates at individual CG sites as well as the Mendelian segregation and fixation of de novo epimutations. CG sites within gbM genes show increased divergence compared to genome-wide divergence rates. Difference of divergence rates significant at $p < 0.001$. **(c)** Observed methylation levels in gbM genes stay approximately constant over generation time (pink bars), which implies that gbM is in steady-state. Methylation levels in gbM genes can be accurately predicted using only estimates of the α and β rates (dark bar; see also Box 1).

A MET1-centered molecular model of the gbM gain and loss rates

One of the current challenges is to establish a link between statistical insights into epimutation rates and the biology of DNA methylation pathways. These rates must, in some way, reflect the accuracy of the maintenance methyltransferases that act on CG dinucleotides. A simple molecular model proposes that the stochastic loss of CG methylation occurs due to the imperfect enzymatic activity of METHYLTRANSFERASE 1 (MET1), which is the primary methyltransferase that targets gbM genes. During DNA replication, hemi-methylated CG sites are detected by the VARIANT IN METHYLATION protein family, which recruits MET1

Box 1. Epimutation rates define steady-state gbM levels

Let **S** be the set of CG dinucleotides corresponding to gene body methylated (gbM) genes. An empirical measure of the methylation level (M) on **S** at any time t is given by:

$$M(S,t) = \frac{N_1(t)+0.5\times N_2(t)}{N},$$

where N_1 and N_2 are the number of homozygous mCG|mCG and heterozygous mCG|CG epigenotypes in set **S**, respectively, and N is the total number of CG sites that are interrogated. Now, let α_S=Pr(CG→mCG) and β_S=Pr(mCG → CG) be the rates of stochastic gain and loss of methylation per CG per unit time, respectively. Using the theoretical model developed by van der Graaf et al. (2015) and Shahryary et al. (2021) an estimator of steady-state M(S, $t \to \infty$) in a selfing system can be given solely in terms of α_S and β_S:

$$\widehat{M}(S,t_\infty) = \widehat{\pi_1}+0.5\widehat{\pi_2}$$

where

$$\widehat{\pi_1} = \frac{\widehat{\alpha_S}((1-\widehat{\alpha_S})^2-(1-\widehat{\beta_S})^2-1)}{(\widehat{\alpha_S}+\widehat{\beta_S})((\widehat{\alpha_S}+\widehat{\beta_S}-1)^2-2)} \text{and } \widehat{\pi_2} = \frac{4\widehat{\alpha_S}\widehat{\beta_S}(\widehat{\alpha_S}+\widehat{\beta_S}-2)}{(\widehat{\alpha_S}+\widehat{\beta_S})((\widehat{\alpha_S}+\widehat{\beta_s}-1)^2-2)}.$$

Several studies have shown that knowledge of α_S and β_S is sufficient to not only predict steady-state methylation (**Fig. 1c**), but also to accurately describe the dynamics of how epimutations accumulate in plant genomes over generations in experimental settings (**Fig. 1b**).

However, steady-state methylation levels vary within gbM genes (**Fig. 3a**). This implies that set **S** can be further partitioned into k subsets S_1, S_2, …, S_k, with corresponding gain rates $\alpha_{S1}, \alpha_{S2}, ..., \alpha_{Sk}$ and loss rates $\beta_{S1}, \beta_{S2}, ..., \beta_{Sk}$. An emerging goal is to define a feature vector based on DNA sequence or chromatin information that can uniquely define all k subsets. Such features could include enrichment information about histone variants or histone modifications, CG density, distance to TSS, etc.

If we let **X** be such a feature vector, and P(**X**) be a set partition function, then we can represent this function as: P(X) : **S** → {S_1, S_2, ..., S_k}, where P(**X**) maps the feature vector **X** to a set partition of the set **S** into k non-empty disjoint subsets. There are fundamental biological questions that can be formulated in the context of this approach. These include: What biological features are important to define vector **X**? How many distinct subsets k are there? Through which molecular mechanisms do the features in **X** affect epimutation rates α_{Si} and β_{Si} *(i = 1, …, k)*? Answering these questions requires a quantitative modeling approach coupled with multi-omics measurements in the context of time-series experiments.

to deposit CG methylation on the newly synthesized strand by way of "template copying" [36,37]. Any accidental failures of MET1 to carry out this catalytic step can lead to permanent methylation losses in daughter cells and their descendant lineages, including the germline (Figure 2a). It has been suggested that MET1 can also have occasional *de novo* activity [38]. This could explain the stochastic gains of CG methylation within gbM genes (Figure 2a). Support for this possibility comes from the fact that similar *de novo* functions have been reported for DNA (cytosine-5)-methyltransferase 1 (DNMT1) [39], the mammalian homolog of MET1; although in mammals, this seems to be restricted primarily to retrotransposons [40].

The distribution of steady-state gbM is mainly driven by variation in the gain rate

However, there are several lines of evidence that a model solely focused on MET1 enzymatic fidelity is insufficient to fully explain epimutational processes within gbM genes. For instance, steady-state levels of gbM are not uniformly distributed along the body of genes. Instead, they display a characteristic decline toward both the 5′ and 3′ ends, with a peak in the center [10,41,42] (Figure 3a). Thus, the relationship between the gain rate (α) and loss rate (β) is modulated by some factors that affect MET1 activity, either directly or indirectly. These factors would need to be identified and incorporated into any epimutational model of gbM. An important aspect of this is to understand whether the intra-genic variation in steady-state gbM levels is more influenced by variation in the gain rate (α) or the loss rate (β). To obtain the first insights into this, we created meta-profiles of these rates within sliding windows along gbM genes (Figure 3b, Supplemental Information). Our profiles reveal an approximately constant loss rate in gene bodies compared to up- and downstream regions, while the gain rate is increased and peaks toward the 3' end. Hence, the gain rate (α) is the main driver of intra-genic variation in steady-state gbM levels. Knowledge of its molecular determinants is, therefore, crucial for understanding gbM dynamics.

Local chromatin features correlate with gain rate variation along gbM genes

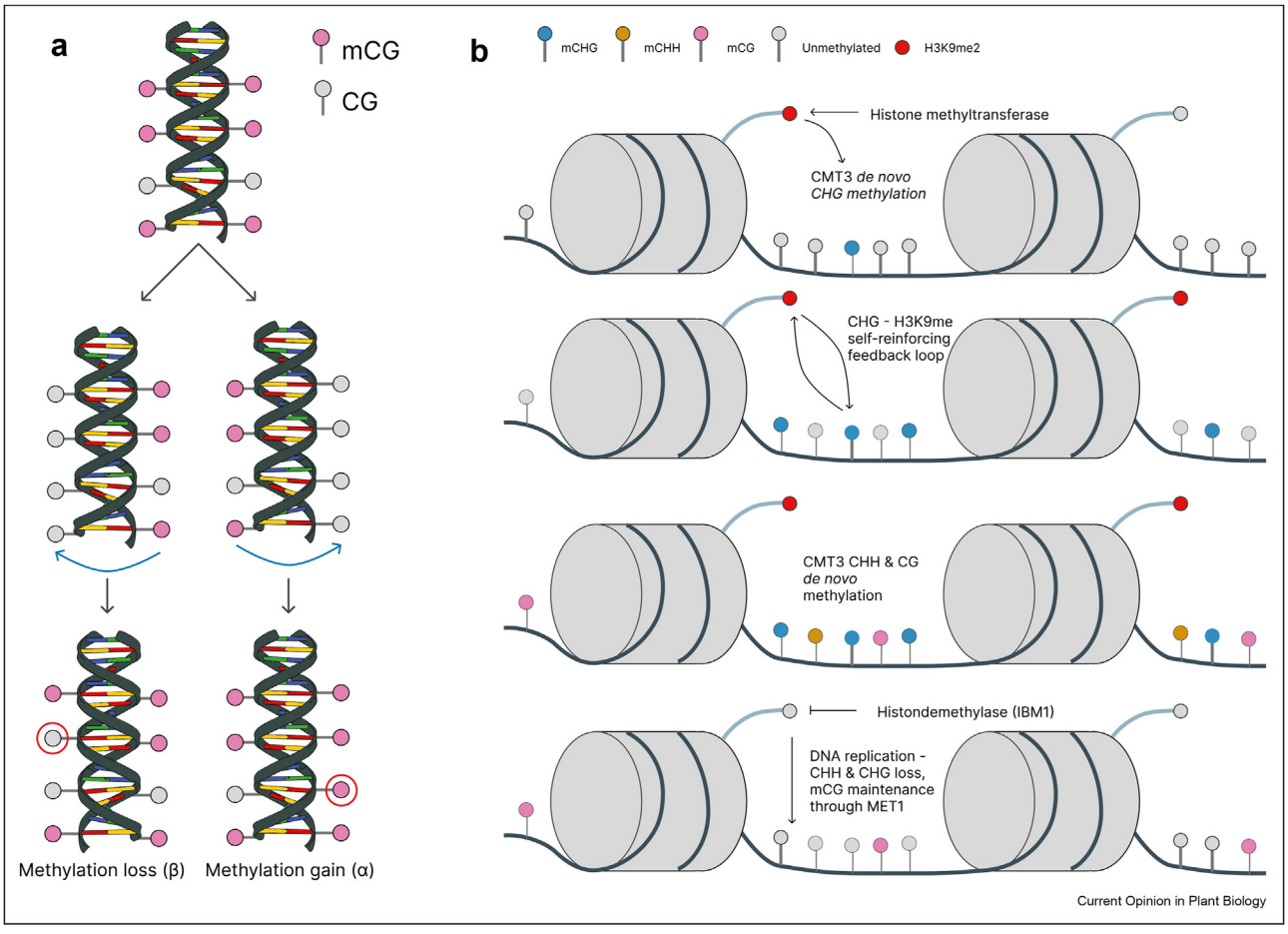
In a first step in this direction, Briffa et al. [29] proposed that MET1 *de novo* activity (i.e. the gain rate) is linked to the distribution of CG density within gbM genes

(Figure 3c). The authors formalized a "cooperative model" in which stochastic gains of methylation at a given CG dinucleotide are more likely to occur if a nearby CG site is also methylated. But this sequence-based model has limitations as CG density is only moderately correlated with gain rate variation within gbM genes (Figure 3d), and a poor predictor at the 5' end where CG density is actually highest [29]. An alternative model is that gain rate variation is mainly determined at the chromatin level [24].

To explore this possibility, we examined the distribution of the histone variant H2A.Z within the body of gbM genes. H2A.Z is typically enriched at the 5′ end of actively transcribed genes and is mutually exclusive with DNA methylation [43] (Figure 3c), and thus highly relevant for understanding patterns of steady-state gbM. Our analysis shows that H2A.Z is more strongly associated with gain rate variation than CG density (Figure 3e). In fact, we speculate that CG density may not even be necessary in a predictive model once H2A.Z enrichment is accounted for, although this would need to be confirmed using statistical model selection procedures.

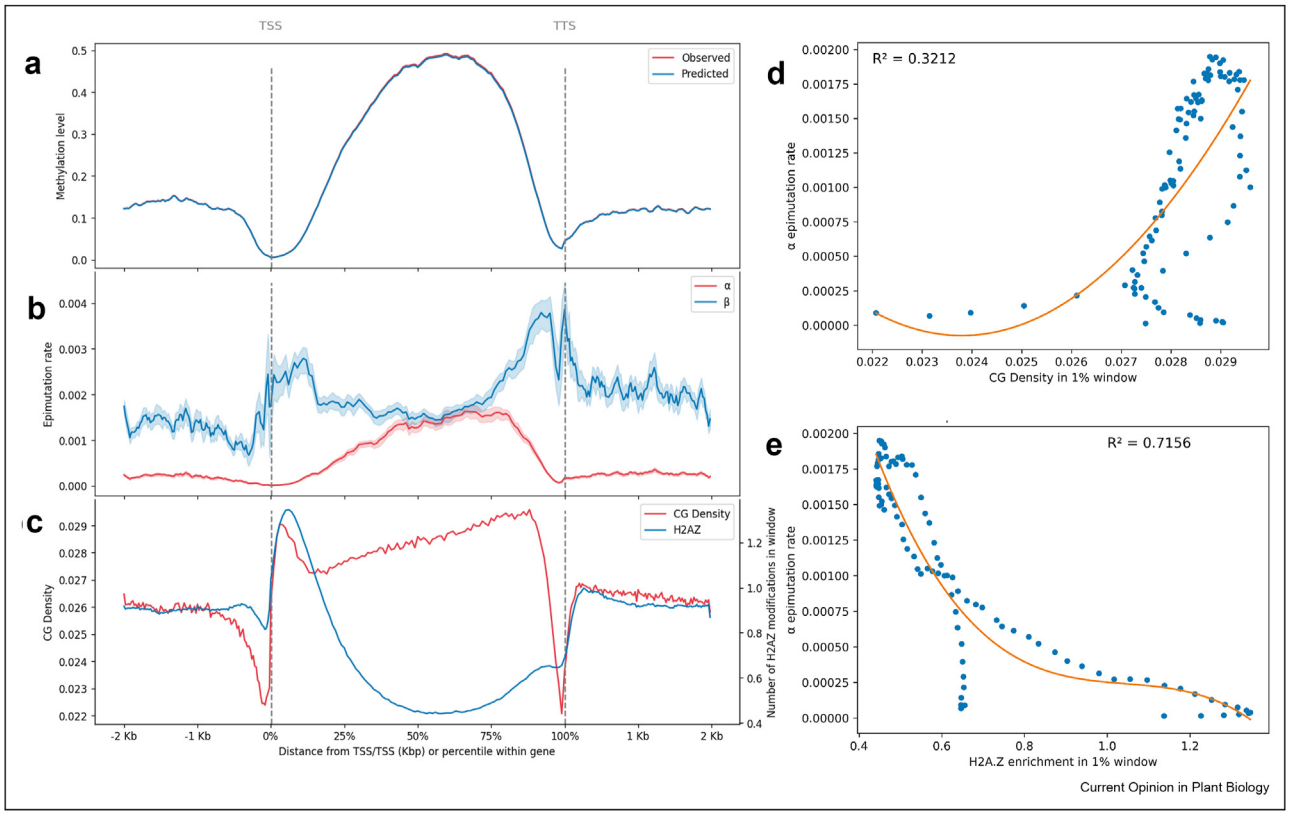
Curiously, first-generation mutants lacking H2A.Z do not display strong shifts in steady-state gbM [43], nor is the distribution of H2A.Z visibly affected in plants

Figure 2



Molecular models for *de novo* CG epimutations in gbM genes. (**a**) A MET1-only model for *de novo* epimutations in gbM genes: for methylated CG dinucleotides, both DNA strands carry the information required to restore the original methylation on the newly synthesized, unmethylated strand. During DNA replication, MET1 is recruited to hemimethylated CG sites to deposit CG methylation on the newly synthesized strand by way of template copying. Any accidental failures of MET1 to carry out this catalytic step can lead to permanent methylation losses in daughter cells (i.e. the rate β). MET1 can also erroneously deposit methylation on untemplated CG dinucleotides, which results in gain of methylation (i.e. the rate α). (**b**) A CMT3-dependent model for *de novo* epimutations in gbM genes: heterochromatin-associated H3K9me2 is accidentally established in gbM genes. As H3K9me2 serves as a substrate for CMT3, it can induce *de novo* methylation at CHG sites, leading to a self-reinforcing mCHG-H3K9me2 feedback loop. The presence of mCHG or H3K9me2 stimulates MET1 *de novo* activity in close proximity. While H3K9me2, and thus also CHG methylation, is efficiently cleared by INCREASED IN BONSAI EXPRESSION 1 (IBM1) from the body of gbM genes [55–57], any accidental gain in CG methylation will be maintained by MET1 upon cell divisions.

Figure 3



Metaplots of CG methylation levels, CG density, H2A.Z enrichment, α and β rates along A. thaliana gbM genes. (a) Metaplots of predicted and observed steady state CG methylation levels in a core set of ~5000 gbM genes from Zhang et al. [52]. GbM genes were separated into 100 sliding windows, spanning 5% of the length of each gene, including 2kb upstream and downstream of the gene. The α and β rates as well as 95% confidence intervals of each window were estimated using AlphaBeta [35]. **(b)** Metaprofiles of the estimated α and β rates along gbM genes, colored areas shows 95% confidence intervals. While intra-genic β stays approximately constant compared to upstream and downstream sequences, α rates deviate from inter-genic levels and peak towards the 3' end of gbM genes. The abrupt peaks in the β rate close to the transcription termination site result from a small sample size of CGs in this region, which is less than 1% of the average intragenic sample size. **(c)** Metaplots of CG density and the H2A.Z histone variant along 1% windows of gbM genes. Both CG density and H2A.Z enrichment deviate from inter-genic baseline levels. H2A.Z data from GSE128434. **(d)** Dependence of intra-genic α on CG density analyzed by linear regression, third degree polynomial line of best fit. **(e)** Dependence of intragenic α on H2A.Z enrichment analyzed by linear regression, third degree polynomial line of best fit.

lacking gbM [6,44]. This would at first suggest that H2A.Z has no causal role in gbM maintenance dynamics. However, a recent multi-generational analysis of H2A.Z mutants revealed that stochastic gains in CG methylation (i.e. the gain rate) are increased several folds, particularly in those regions of gbM genes that are typically unmethylated in wild-type plants (e.g. at the 5' end) [29]. Hence, H2A.Z is a major component of chromatin that appears to shape gbM dynamics by suppressing the gain rate in specific regions, either directly or indirectly. This effect is subtle, and its impact on steady gbM only becomes visible over longer timescales. Indeed, with the estimated gain rate changes observed in H2A.Z mutants, it would take ~2000 generations for gbM regions that are targeted by H2A.Z to reach their new steady state [29].

Stochasticity in gbM is responsive to heterochromatin in trans

There is additional evidence to suggest that MET1 does not act alone in causing epimutations in gbM genes. Recent work has shown that altering DNA methylation in other regions of the genome can also affect gbM [24,29,45–50]. For instance, mutations in CHROMOMETHYLASE 3 (CMT3) or the plant-specific histone methyltransferases SU(VAR)4/5/6, which are responsible for maintaining CHG methylation [51], lead to slightly increased gbM levels and a significantly decreased CG methylation gain and loss rates [24] (Figs. S1a–b). The fact that the reduction of CHG methylation in these mutants occurs primarily in heterochromatin indicates that gbM maintenance dynamics are somehow responsive to the heterochromatic status of cells [24,52].

The link between CMT3 and gbM comes from earlier comparative (epi)genomic studies of angiosperms, which revealed that several species that had lost CMT3 in their evolutionary past also lost gbM [2,12]. Based on these observations, Bewick et al. [6] suggested that gbM is simply a by-product of heterochromatin maintenance. They expressed this idea in a molecular model in which heterochromatin-associated histone H3 lysine 9 dimethylation (H3K9me2) is occasionally established in gbM genes by accident [53], or by some unknown nuclear process. As H3K9me2 serves as a substrate for CMT3 [54], it facilitates transient CHG methylation in gbM genes. The presence of mCHG can have knock-on effects that occasionally lead to the deposition of methylation at neighboring CG sites (Figure 2b). This latter process is not fully understood [6], but it is possible that the presence of mCHG or H3K9me2 stimulates proximal *de novo* MET1 activity. While H3K9me2, and thus also CHG methylation, is efficiently cleared by INCREASED IN BONSAI EXPRESSION 1 (IBM1) from the body of gbM genes [55–57], any accidental gain in CG methylation will be maintained by MET1 upon cell divisions. The fact that IBM1 is not correctly spliced in plants lacking CMT3 and SUV4/5/6 [58] further suggests that the trans effect of heterochromatin could include altered activity by IBM1.

That this stochastic model of the gain rate (i.e. α) does not lead to complete CG hypermethylation in gbM genes over time is likely due to the simultaneous failure of MET1 to maintain CG methylation (i.e. the loss rate (β)). The significant reduction in the gain rate seen in mutation accumulation lines that lack CMT3 and SUV4/5/6 is consistent with this model (Figs. S1a–b). However, those lines also display a reduced loss rate and slightly increased gbM levels, which does not exactly match the DNA methylation signatures seen in plant species that have lost CMT3 in their evolutionary past. This discrepancy highlights the general difficulty in trying to reconcile evolutionary with experimental timescales [33], but points to yet unknown components that regulate gbM dynamics.

Reconciling long-term evolutionary with short-term laboratory dynamics in gbM

From our perspective, a central question is how to connect epimutational processes that occur at single CGs when gbM is maintained at steady state with the drastic steady-state transitions (i.e. complete loss or gain of gbM in specific genes) that occasionally occur over evolutionary timescales. Although gbM is conserved, variation in gbM status does exist between closely related species [2,15] and even among individuals of the same species [52,59]. This variation implies that genes have the ability to transition from an unmethylated state (UM) to a gbM state [52] (Figure 1a). But why some UM genes are targeted for gbM in the first place and not others remain largely unknown. An important genic feature that has been associated with increased transition probabilities is gene length and an enrichment of trinucleotide context CWG (W = A or T), a CMT3 substrate [15,52]. Under what circumstance steady-state shifts are actually triggered and how new steady-states are stabilized, subsequently, remains unclear.

One way to identify pathways that mediate transitions in steady-state gbM is through genetic mapping studies in natural populations, where genic methylation is treated as a complex molecular trait [29,46,59–63]. Briffa et al. [29], for instance, demonstrated that variation in global gbM levels measured among ~1000 *Arabidopsis* accessions are associated with genetic polymorphisms in HTA9, one of three genes encoding the H2A.Z protein in *Arabidopsis* [64]. Accessions carrying the minor HTA9-G allele exhibit higher global gbM and lower expression of HTA9. This is consistent with the elevated gain rates and reduced loss rates seen in H2A.Z mutants and suggests that the lower expression of HTA9 is likely responsible for high gbM in HTA9-G accessions. To test if differential H2A.Z enrichment truly mediates these associations, genome-wide profiling of H2A.Z among these associations would be necessary.

However, the type of associations that can be uncovered using genetic mapping approaches of gbM appears to depend strongly on how gbM is defined as a target trait in the first place. For instance, instead of using global gbM levels, Zhang et al. [52] performed GWAS on the number of gbM genes found among different accessions, a definition that is more consistent with steady-state transitions in gbM. They uncovered an association with polymorphisms at a locus containing gene RUG1, a putative regulator of chromosome condensation, and a H1 linker histone. It is still unclear how these candidate loci affect epimutational processes, but they do provide useful leads for building a more comprehensive molecular model that would bridge mechanistic insights into gbM maintenance dynamics with its long-term evolution.

Discussion

Studying steady-state gbM from the perspective of a stochastic process can provide novel insights into its molecular mechanisms and evolution. Statistical estimates of CG methylation gain and loss rates hold potentially useful information about the enzymatic activity of methyltransferases targeting gbM genes. We have shown here that intra-genic variation in steady-state gbM levels is mainly the result of variation in the gain rate along the body of genes. Whether the gain rate is mainly shaped by the—as yet poorly characterized—*de novo* activity of MET1 modulated by the presence of H2A.Z, or by CMT3-mediated mechanisms involving stochastic mis-integration of H3K9me2, remains unclear.

The answer is probably multifactorial, with contributions from both. Regardless, intra-genic variation in the gain rate can be well predicted from local chromatin features, and epimutational processes are sensitive to the heterochromatic state of cells. This observation places chromatin organization at the center stage of any molecular model attempting to explain the stochasticity in gbM in plants.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have re-analyzed public data.

Acknowledgements

FJ acknowledges support from the Deutsche Forschungsgemeinschaft (DFG). We thank Robert J. Schmitz for feedback. We thank Zhilin Zhang for his help with data preparation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pbi.2023.102436.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

1. Tran RK, Henikoff JG, Zilberman D, Ditt RF, Jacobsen SE, Henikoff S: **DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes**. *Curr Biol* 2005, **15**: 154–159.

2. Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, *et al.*: **Widespread natural variation of DNA methylation within angiosperms**. *Genome Biol* 2016, **17**:194.

3. Takuno S, Ran J-H, Gaut BS: **Evolutionary patterns of genic DNA methylation vary across land plants**. *Native Plants* 2016, **2**, 15222.

4. Bewick AJ, Niederhuth CE, Ji L, Rohr NA, Griffin PT, Leebens-Mack J, Schmitz RJ: **The evolution of CHROMOMETHYLASES and gene body DNA methylation in plants**. *Genome Biol* 2017, **18**:65.

5. Marchant DB, Chen G, Cai S, Chen F, Schafran P, Jenkins J, Shu S, Plott C, Webber J, Lovell JT, *et al.*: **Dynamic genome evolution in a model fern**. *Native Plants* 2022, **8**:1038–1051.

6. Bewick AJ, Schmitz RJ: **Gene body DNA methylation in plants**. *Curr Opin Plant Biol* 2017, **36**:103–110.

7. Horvath R, Laenen B, Takuno S, Slotte T: **Single-cell expression noise and gene-body methylation in Arabidopsis thaliana**. *Heredity* 2019, **123**:81–91.

8. 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**:61–69.

9. Vaughn MW, Tanurdžić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, *et al.*: **Epigenetic natural variation in Arabidopsis thaliana**. *PLoS Biol* 2007, **5**:e174.

10. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW-L, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, *et al.*: **Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis**. *Cell* 2006, **126**: 1189–1201.

11. Muyle AM, Seymour DK, Lv Y, Huettel B, Gaut BS: **Gene body methylation in plants: mechanisms, functions, and important implications for understanding evolutionary processes**. *Genome Biol Evol* 2022:14.

12. Kiefer C, Willing E-M, Jiao W-B, Sun H, Piednoël M, Hümann U, Hartwig B, Koch MA, Schneeberger K: **Interspecies association mapping links reduced CG to TG substitution rates to the loss of gene-body methylation**. *Nature Plants* 2019, **5**: 846–855.

13. Seymour DK, Gaut BS: **Phylogenetic shifts in gene body methylation correlate with gene expression and reflect trait conservation**. *Mol Biol Evol* 2020, **37**:31–43.

14. Bewick AJ, Zhang Y, Wendte JM, Zhang X, Schmitz RJ: **Evolutionary and experimental loss of gene body methylation and its consequence to gene expression**. *G3: Genes, Genomes, Genetics* 2019, https://doi.org/10.1534/g3.119.400365.

15. Takuno S, Seymour DK, Gaut BS: **The evolutionary dynamics of orthologs that shift in gene body methylation between arabidopsis species**. *Mol Biol Evol* 2017, **34**:1479–1491.

16. Takuno S, Gaut BS: **Body-methylated genes in Arabidopsis thaliana are functionally important and evolve slowly**. *Mol Biol Evol* 2012, **29**:219–227.

17. Zemach A, McDaniel IE, Silva P, Zilberman D: **Genome-wide evolutionary analysis of eukaryotic DNA methylation**. *Science* 2010, **328**:916–919.

18. Feng S, Cokus SJ, Zhang X, Chen P-Y, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, *et al.*: **Conservation and divergence of methylation patterning in plants and animals**. In *Proceedings of the National Academy of Sciences*, **107**; 2010:8689–8694.

19. Takuno S, Gaut BS: **Gene body methylation is conserved between plant orthologs and is of evolutionary consequence**. In *Proceedings of the National Academy of Sciences*, **110**; 2013: 1797–1802.

20. Zilberman D: **An evolutionary case for functional gene body methylation in plants and animals**. *Genome Biol* 2017, **18**.

21. Muyle A, Gaut BS: **Loss of gene body methylation in eutrema salsugineum is associated with reduced gene expression**. *Mol Biol Evol* 2019, **36**:155–158.

22. Wang H, Beyene G, Zhai J, Feng S, Fahlgren N, Taylor NJ, Bart R, Carrington JC, Jacobsen SE, Ausin I: **CG gene body DNA methylation changes and evolution of duplicated genes in cassava**. In *Proceedings of the National Academy of Sciences*, **112**; 2015:13729–13734.

23. Xu G, Lyu J, Li Q, Liu H, Wang D, Zhang M, Springer NM, Ross-Ibarra J, Yang J: **Evolutionary and functional genomics of DNA methylation in maize domestication and improvement**. *Nat Commun* 2020, **11**:5539.

24. * Hazarika RR, Serra M, Zhang Z, Zhang Y, Schmitz RJ, Johannes F: **Molecular properties of epimutation hotspots**. *Native Plants* 2022, **8**:146–156.
This study shows that gbM genes harbor CG epimutation hotspots in *A. thaliana* and that CG epimutation rates within gbM genes are sensitive to heterochromatin methylation in trans.

25. Yao N, Zhang Z, Yu L, Hazarika R, Yu C, Jang H, Smith LM, Ton J, Liu L, Stachowicz J, *et al.*: **An evolutionary epigenetic clock in plants**. *bioRxiv* 2023, https://doi.org/10.1101/2023.03.15.532766.

26. Hofmeister BT, Lee K, Rohr NA, Hall DW, Schmitz RJ: **Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation**. *Genome Biol* 2017, **18**:155.

27. Picard CL, Gehring M: **Proximal methylation features associated with nonrandom changes in gene body methylation**. *Genome Biol* 2017, **18**:73.

28. Denkena J, Johannes F, Colomé-Tatché M: **Region-level epimutation rates in Arabidopsis thaliana**. *Heredity* 2021, **127**: 190–202.

29. * Briffa A, Hollwey E, Shahzad Z, Moore JD, Lyons DB, Howard M, Zilberman D: **Unified establishment and epigenetic inheritance of DNA methylation through cooperative MET1 activity**. *bioRxiv* 2022, https://doi.org/10.1101/2022.09.12.507517.
This study proposes that stochastic gains in CG methylation within gbM genes is the result of MET1 *de novo* activity, which works in conjunction with local CG density.

30. Johannes F, Schmitz RJ: **Spontaneous epimutations in plants**. *New Phytol* 2019, **221**:1253–1259.

31. Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, Weigel D: **Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome**. *Nature* 2011, **480**:245–249.

32. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR: **Transgenerational epigenetic instability is a source of novel methylation variants**. *Science* 2011, **334**:369–373.

33. Vidalis A, Živković D, Wardenaar R, Roquis D, Tellier A, Johannes F: **Methylome evolution in plants**. *Genome Biol* 2016, **17**:264.

34. van der Graaf A, Wardenaar R, Neumann DA, Taudt A, Shaw RG, Jansen RC, Schmitz RJ, Colomé-Tatché M, Johannes F: **Rate, spectrum, and evolutionary dynamics of spontaneous epimutations**. In *Proceedings of the National Academy of Sciences*, **112**; 2015:6676–6681.

35. Shahryary Y, Symeonidi A, Hazarika RR, Denkena J, Mubeen T, Hofmeister B, van Gurp T, Colomé-Tatché M, Verhoeven KJF, Tuskan G, *et al.*: **AlphaBeta: computational inference of epimutation rates and spectra from high-throughput DNA methylation data in plants**. *Genome Biol* 2020, **21**:1–22.

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

37. Woo HR, Pontes O, Pikaard CS, Richards EJ: **VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization**. *Genes Dev* 2007, **21**:267–277.

38. Zubko E, Gentry M, Kunova A, Meyer P: **De novo DNA methylation activity of methyltransferase 1 (MET1) partially restores body methylation in Arabidopsis thaliana**. *Plant J* 2012, **71**:1029–1037.

39. Jeltsch A, Adam S, Dukatz M, Emperle M, Bashtrykov P: **Deep enzymology studies on DNA methyltransferases reveal novel connections between flanking sequences and enzyme activity**. *J Mol Biol* 2021, **433**, 167186.

40. * Haggerty C, Kretzmer H, Riemenschneider C, Kumar AS, Mattei AL, Bailly N, Gottfreund J, Giesselmann P, Weigert R, Brändl B, *et al.*: **Dnmt1 has de novo activity targeted to transposable elements**. *Nat Struct Mol Biol* 2021, **28**:594–603.
This animal study demonstrates that MET1 homolog Dnmt1 displays *de novo* methylation activity in vitro and in vivo with specific retrotransposon targeting.

41. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR: **Highly integrated single-base resolution maps of the epigenome in Arabidopsis**. *Cell* 2008, **133**: 523–536.

42. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE: **Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning**. *Nature* 2008, **452**:215–219.

43. Coleman-Derr D, Zilberman D: **Deposition of histone variant H2A.Z within gene bodies regulates responsive genes**. *PLoS Genet* 2012, **8**, e1002988.

44. Bewick, Adam J, Lexiang Ji, Chad E. Niederhuth, Eva-Maria Willing, Brigitte T. Hofmeister, Xiuling Shi, Li Wang, *et al.*: **"On the origin and evolutionary consequences of gene body DNA methylation"**. *Proc Natl Acad Sci U S A* 2016, **113**: 9111–9116.

45. Kakoulidou I, Piecyk RS, Meyer RC, Kuhlmann M, Gutjahr C, Altmann T, Johannes F: **Parental pericentromeric methylation status drives methylome remodelling and heterosis in epigenetic hybrids**. *bioRxiv* 2022, https://doi.org/10.1101/2022.09.29.510107.

46. * Zhang Y, Jang H, Xiao R, Kakoulidou I, Piecyk RS, Johannes F, Schmitz RJ: **Heterochromatin is a quantitative trait associated with spontaneous epiallele formation**. *Nat Commun* 2021, **12**: 6958.
This study demonstrates that CHG methylation in genes co-varies with the amount of heterochromatin methylation. This covariation is particularly abundant for gbM genes.

47. To TK, Yamasaki C, Oda S, Tominaga S, Kobayashi A, Tarutani Y, Kakutani T: **Local and global crosstalk among heterochromatin marks drives DNA methylome patterning in Arabidopsis**. *Nat Commun* 2022, **13**:861.

48. 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**:352–364.

49. Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE: **Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis**. *Nat Struct Mol Biol* 2014, **21**: 64–72.

50. Ito T, Tarutani Y, To TK, Kassam M, Duvernois-Berthet E, Cortijo S, Takashima K, Saze H, Toyoda A, Fujiyama A, *et al.*: **Genome-wide negative feedback drives transgenerational DNA methylation dynamics in arabidopsis**. *PLoS Genet* 2015, **11**, e1005154.

51. Jackson JP, Lindroth AM, Cao X, Jacobsen SE: **Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase**. *Nature* 2002, **416**:556–560.

52. * Zhang Y, Wendte JM, Ji L, Schmitz RJ: **Natural variation in DNA methylation homeostasis and the emergence of epialleles**. In *Proceedings of the National Academy of Sciences*; 2020, https://doi.org/10.1073/pnas.1918172117.
This study identifies gbM genes in *A. thaliana* natural accessions, and uses variation in the number of gbM genes as a molecular trait for genome-wide association mapping. It identifies candidate loci that may contribute to evolutionary shifts in steady-state methylation; that is, genic transitions from an unmethylated state to a gbM state.

53. Wendte JM, Zhang Y, Ji L, Shi X, Hazarika RR, Shahryary Y, Johannes F, Schmitz RJ: **Epimutations are associated with CHROMOMETHYLASE 3-induced de novo DNA methylation**. *Elife* 2019, **8**, e47891.

54. Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A, *et al.*: **Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants**. *Cell* 2012, **151**: 167–180.

55. Saze H, Shiraishi A, Miura A, Kakutani T: **Control of genic DNA methylation by a jmjC domain-containing protein in Arabidopsis thaliana**. *Science* 2008, **319**:462–465.

56. Inagaki S, Miura-Kamio A, Nakamura Y, Lu F, Cui X, Cao X, Kimura H, Saze H, Kakutani T: **Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome**. *EMBO J* 2010, **29**:3496–3506.

57. Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, Kakutani T: **An Arabidopsis jmjC domain protein protects transcribed genes from DNA methylation at CHG sites**. *EMBO J* 2009, **28**:1078–1086.

58. Rigal M, Kevei Z, Pélissier T, Mathieu O: **DNA methylation in an intron of the IBM1 histone demethylase gene stabilizes chromatin modification patterns**. *EMBO J* 2012, **31**: 2981–2993.

59. Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, *et al.*: **Epigenomic diversity in a global collection of Arabidopsis thaliana accessions**. *Cell* 2016, **166**:492–505.

60. Dubin MJ, Zhang P, Meng D, Remigereau M-S, Osborne EJ, Paolo Casale F, Drewe P, Kahles A, Jean G, Vilhjálmsson B,

et al.: **DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation**. *Elife* 2015, **4**, e05255.

61. Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, Alix A, McCosh RB, Chen H, Schork NJ, *et al.*: **Patterns of population epigenomic diversity**. *Nature* 2013, **495**:193–198.

62. Taudt A, Colomé-Tatché M, Johannes F: **Genetic sources of population epigenomic variation**. *Nat Rev Genet* 2016, **17**: 319–332.

63. Pisupati R, Nizhynska V, Mollá Morales A, Nordborg M: **On the causes of gene-body methylation variation in Arabidopsis thaliana**. *PLoS Genet* 2023, **19**, e1010728.

64. Yi H, Sardesai N, Fujinuma T, Chan C-W, Veena, Gelvin SB: **Constitutive expression exposes functional redundancy between the Arabidopsis histone H2A gene HTA1 and other H2A gene family members**. *Plant Cell* 2006, **18**: 1575–1589.