Rate, spectrum, and evolutionary dynamics of spontaneous epimutations

Adriaan van der Graaf[a,1], René Wardenaar[a,1], Drexel A. Neumann[b], Aaron Taudt[c], Ruth G. Shaw[d], Ritsert C. Jansen[a], Robert J. Schmitz[b,2], Maria Colomé-Tatché[c,2], and Frank Johannes[a,2]

[a]Groningen Bioinformatics Centre, University of Groningen, 9747 AG Groningen, The Netherlands; [b]Department of Genetics, University of Georgia, Athens, GA 30602; [c]European Institute for the Biology of Aging, University of Groningen, University Medical Centre Groningen, 9713 AV Groningen, The Netherlands; and [d]Department of Ecology, Evolution and Behavior, University of Minnesota, Minneapolis, MN 55455

Edited by James A. Birchler, University of Missouri–Columbia, Columbia, MO, and approved April 14, 2015 (received for review December 19, 2014)

Stochastic changes in cytosine methylation are a source of heritable epigenetic and phenotypic diversity in plants. Using the model plant *Arabidopsis thaliana*, we derive robust estimates of the rate at which methylation is spontaneously gained (forward epimutation) or lost (backward epimutation) at individual cytosines and construct a comprehensive picture of the epimutation landscape in this species. We demonstrate that the dynamic interplay between forward and backward epimutations is modulated by genomic context and show that subtle contextual differences have profoundly shaped patterns of methylation diversity in *A. thaliana* natural populations over evolutionary timescales. Theoretical arguments indicate that the epimutation rates reported here are high enough to rapidly uncouple genetic from epigenetic variation, but low enough for new epialleles to sustain long-term selection responses. Our results provide new insights into methylome evolution and its population-level consequences.

epigenetics | epimutation | DNA methylation | evolution | *Arabidopsis*

Plant genomes make extensive use of cytosine methylation to control the expression of transposable elements (TEs) and genes (1). Despite its tight regulation, methylation losses or gains at individual cytosines or clusters of cytosines can emerge spontaneously, in an event termed "epimutation" (2, 3). Many examples of segregating epimutations have been documented in experimental and wild populations of plants and in some cases contribute to heritable variation in phenotypes independently of DNA sequence variation (4, 5). These observations have led to much speculation about the role of DNA methylation in plant evolution (6–8), and its potential in breeding programs (9). In the model plant *Arabidopsis thaliana*, spontaneous methylation changes at CG dinucleotides accumulate in a rapid but nonlinear fashion over generations (2, 3, 10), thus pointing to high forward–backward epimutation rates (11). Precise estimates of these rates are necessary to be able to quantify the long-term dynamics of epigenetic variation under laboratory or natural conditions, and to understand the molecular mechanisms that drive methylome evolution (12–14). Here we combine theoretical modeling with high-resolution methylome analysis of multiple independent *A. thaliana* mutation accumulation (MA) lines (15), including measurements of methylation changes in continuous generations, to obtain robust estimates of forward and backward epimutation rates.

Results

We joined whole-genome MethylC-seq (16) data from two earlier MA studies (2, 3) with extensive multigenerational MethylC-seq measurements from three additional MA lines (Fig. 1*A* and *SI Appendix*, Tables S1–S6). The first of these new MA lines (MA1_3) was propagated for 30 generations and includes measurements for 13 (nearly) consecutive generations (Fig. 1*A*). The other two MA lines (MA2_3) were propagated for 17 generations and were measured every four generations on average (Fig. 1*A*). These new data therefore allowed us to track epimutation dynamics over a large number of generations and at high temporal resolution. We constructed base pair-resolution methylation maps for all sequenced individuals (*SI Appendix*). To obtain a measure of genome-wide methylation divergence between any two individuals in a given MA pedigree, we calculated the proportion of differentially methylated cytosines in sequence contexts CG, CHG, and CHH (where H can be any base but G). For these calculations we used a set of consensus cytosines for which all individuals in the pedigrees had coverage of more than three reads (*SI Appendix*). This read coverage cutoff was found to be sufficient for robust downstream analyses (*SI Appendix*, Figs. S1 and S2). Consistent with previous reports (2, 3, 10), genome-wide methylation divergence at CG dinucleotides increased with divergence time in all pedigrees (Fig. 1*B*), but not in sequence contexts CHG and CHH (*SI Appendix*, Fig. S3). This distinction reflects intrinsic differences in the maintenance pathways that target these three contexts (1) and possibly also increased measurement error and cellular heterogeneity for non-CG methylation (*SI Appendix*, Fig. S4).

Neutral Epimutation Model. To quantify CG methylation divergence in the MA lines as a function of divergence time (measured in generations) and forward–backward epimutation rates, we developed a theoretical model similar to those used in the analysis of regular systems of inbreeding (*Materials and Methods* and *SI Appendix*). Briefly, the model assumes that an unmethylated

Significance

Changes in the methylation status of cytosine nucleotides are a source of heritable epigenetic and phenotypic diversity in plants. Here we derive robust estimates of the rate at which cytosine methylation is spontaneously gained (forward epimutation) or lost (backward epimutation) in the genome of the model plant *Arabidopsis thaliana*. We show that the forward–backward dynamics of selectively neutral epimutations have a major impact on methylome evolution and shape genome-wide patterns of methylation diversity among natural populations in this species. The epimutation rates presented here can serve as reference values in future empirical and theoretical population epigenetic studies in plants.

Author contributions: R.J.S., M.C.-T., and F.J. designed research; A.v.d.G., R.W., D.A.N., A.T., R.J.S., M.C.-T., and F.J. performed research; D.A.N., R.G.S., R.C.J., and R.J.S. contributed new reagents/analytic tools; A.v.d.G., R.W., A.T., M.C.-T., and F.J. analyzed data; and M.C.-T. and F.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE64463).

[1]A.v.d.G. and R.W. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: schmitz@uga.edu, m.colome.tatche@umcg.nl, or frank@johanneslab.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1424254112/-/DCSupplemental.

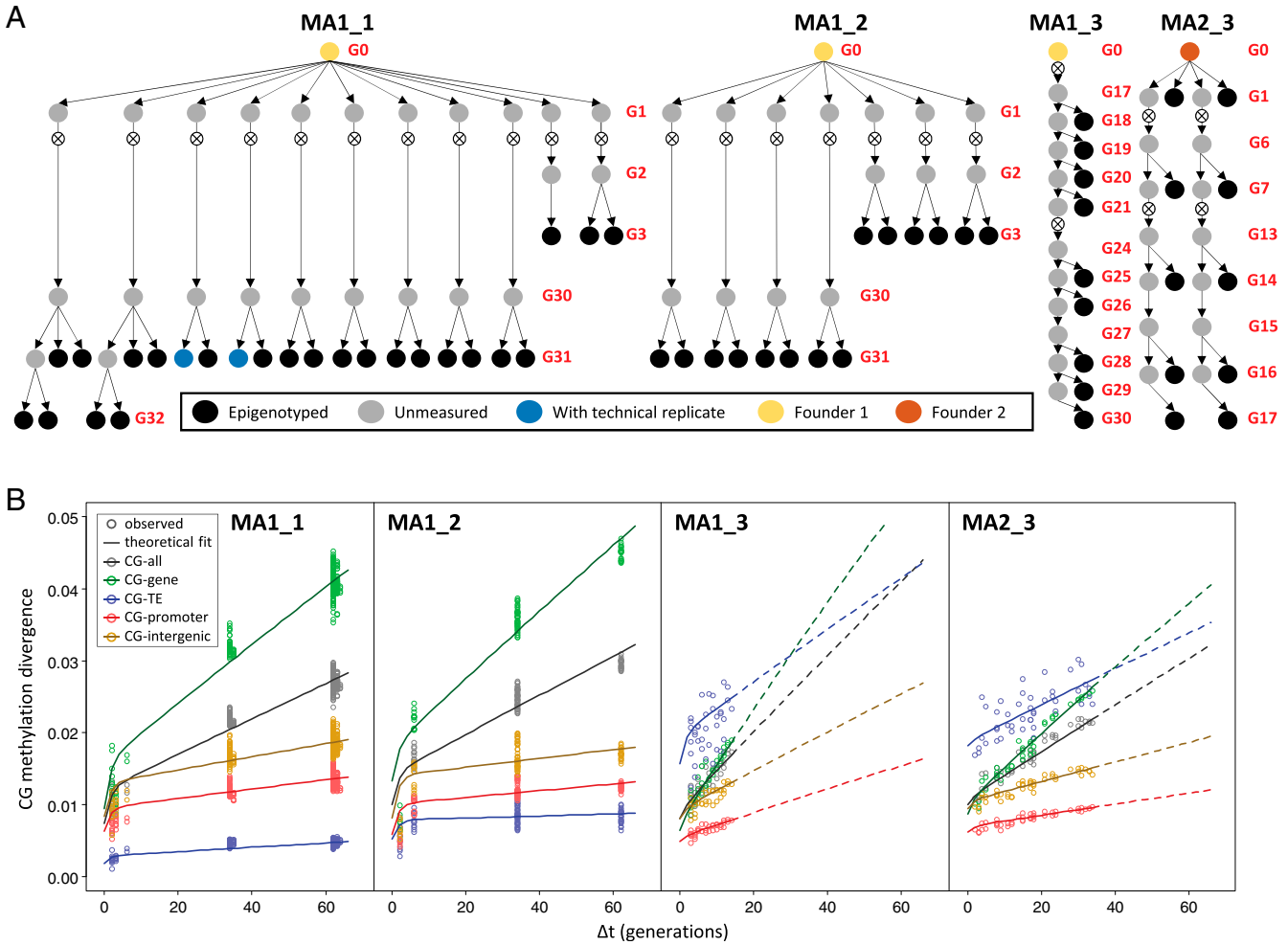

Fig. 1. (*A*) Overview of pedigrees of mutation accumulation lines (MA lines). Red numbers indicate the number of generations from common founder. The MA1_1 and MA1_2 lines were originally created by Shaw et al. (15) and their methylomes were presented in Becker et al. (2) and Schmitz et al. (3), respectively. The MA1_3 and MA2_3 were generated in this study. (*B*) Measured CG methylation divergence (circles) with corresponding theoretical fits (lines) as a function of divergence time between two individuals in a given pedigree (Δt = total divergence time in generations between any two individuals). Dashed lines indicate extrapolation from the fitted model. Divergence values for CG-TE in datasets MA1_3 and MA2_3 were susceptible to low sequencing depth in these experiments and showed increased measurement noise (*SI Appendix*). Model-based analysis of all MA pedigrees revealed that the highly nonlinear divergence until generation eight is due to the fixation of segregating epi-heterozygote founder loci (*SI Appendix*), rather than the result of recurrent cycles of forward and backward epimutations as previously suggested (11).

cytosine (c^u) can become methylated (c^m) with probability α and likewise a methylated cytosine can become unmethylated with probability β. We arbitrarily define α as the forward and β as the backward epimutation rate per generation per haploid methylome. Transitions of diploid epigenotypes (c^mc^m, c^uc^u, or c^mc^u) from one generation to the next are modeled through a transition matrix where the elements of the matrix are determined by the Mendelian segregation of epialleles (c^m or c^u) and the rates α and β. Consistent with the MA experimental design, selection on epigenotypes during inbreeding is assumed to be absent, so that the cumulative divergence among lines is driven solely by neutral epigenetic drift. Estimates for the unknown epimutation rates are obtained by fitting our model to the CG methylation divergence data of each MA pedigree separately (*Materials and Methods*).

Estimates of Global CG Epimutation Rates. As shown in Fig. 1*B*, our model provides an excellent fit to the data, which suggests that the observed divergence patterns among the MA lines are largely the result of the transgenerational accumulation of selectively neutral epimutations. Model-based estimates for the forward and backward CG epimutation rates (CG-all) were $2.56 \cdot 10^{-4}$ and $6.30 \cdot 10^{-4}$, respectively (Table 1 and *SI Appendix*, Table S7). These estimates are similar to the value provided by Schmitz et al. (3) ($4.46 \cdot 10^{-4}$) but illustrate that methylation loss at CG dinucleotides is globally three times as likely as methylation gain. The ratio of loss to gain (β/α), also known as the mutational bias parameter, is an important quantity: It determines the CG methylation content of the *A. thaliana* genome over evolutionary timescales. Assuming that the *A. thaliana* methylome is at equilibrium, the estimated CG forward–backward epimutation rates imply that—in the absence of selection or gene conversion—about 30% of all CG sites should be methylated and about 70% are unmethylated, which is consistent with actual measurements (16, 17).

Estimates of Annotation-Specific CG Epimutation Rates. We examined the extent to which CG epimutation rates depend on genomic context. To do this, we separated all CGs according to annotation (gene bodies, promoters, TEs, and intergenic regions; see *SI Appendix*). Although annotation-specific CG epimutation rates were approximately within the same order of magnitude (Table 1 and *SI Appendix*, Table S7), subtle differences in these rates had a substantial impact on differential divergence of CG methylation

Table 1. Estimates of forward and backward epimutation rates

Context	α	Range (α)		β	Range (β)		β/α	Range (β/α)	
CG-all	$2.56 \cdot 10^{-4}$	$2.08 \cdot 10^{-4}$	$3.69 \cdot 10^{-4}$	$6.30 \cdot 10^{-4}$	$3.23 \cdot 10^{-4}$	$1.13 \cdot 10^{-3}$	2.36	1.55	3.24
CG-gene	$3.48 \cdot 10^{-4}$	$2.77 \cdot 10^{-4}$	$4.87 \cdot 10^{-4}$	$1.47 \cdot 10^{-3}$	$9.46 \cdot 10^{-4}$	$2.45 \cdot 10^{-3}$	4.24	2.84	5.10
CG-TE*	$3.24 \cdot 10^{-4}$	$1.68 \cdot 10^{-4}$	$4.80 \cdot 10^{-4}$	$1.20 \cdot 10^{-5}$	$7.76 \cdot 10^{-6}$	$1.62 \cdot 10^{-5}$	0.040	0.034	0.046
CG-promoter	$5.17 \cdot 10^{-5}$	$2.92 \cdot 10^{-5}$	$9.33 \cdot 10^{-5}$	$5.88 \cdot 10^{-4}$	$1.33 \cdot 10^{-4}$	$1.40 \cdot 10^{-3}$	11.4	4.16	15.08
CG-intergenic	$1.15 \cdot 10^{-4}$	$6.13 \cdot 10^{-5}$	$1.70 \cdot 10^{-4}$	$3.25 \cdot 10^{-4}$	$6.36 \cdot 10^{-5}$	$7.69 \cdot 10^{-4}$	2.83	0.47	4.80

We assume that an unmethylated cytosine (c^u) can become methylated (c^m) with probability α, and likewise a methylated cytosine can become unmethylated with probability β. We arbitrarily define α as the forward and β as the backward epimutation rate per generation per haploid methylome. Shown are model-based estimates for α and β as an average of the MA1_1, MA1_2, MA1_3, and MA2_3 datasets, as well as the range of these estimates across datasets (range). The asterisk indicates that the average estimate was based only on the MA1_1 and the MA1_2 data (*SI Appendix*). These estimates can be considered robust, because the different MA pedigrees varied considerably in terms of plant material, growth conditions, and sequencing approach (*SI Appendix*, Table S1).

across annotation categories (Fig. 1*B*). The highest combined forward and backward rates were found for CGs in gene bodies (CG-gene), which were $3.48 \cdot 10^{-4}$ and $1.47 \cdot 10^{-3}$, respectively (Table 1 and *SI Appendix*, Table S7). By contrast, the lowest rates were found for CGs in TEs (CG-TEs, forward: $3.24 \cdot 10^{-4}$ and backward: $1.20 \cdot 10^{-5}$). As a result of these low epimutation rates, methylation divergence for CG-TEs was much less pronounced (Fig. 1*B*), resembling the divergence patterns seen for CHG and CHH contexts (*SI Appendix*, Fig. S3). This observation suggests that CG-TEs come under the influence of silencing pathways that primarily target neighboring CHHs and CHGs (18–20). Indeed, CG-TE was the only annotation category in which the ratio of backward to forward epimutation rates was less than unity (Table 1 and *SI Appendix*, Table S7), which implies that gain of methylation is strongly favored over methylation loss.

Genome Architecture and Chromatin Environment Predict CG Methylation Divergence Patterns Along Chromosomes. Because CG epimutation rates are annotation-specific, we predicted that methylation divergence closely tracks annotation density along chromosomes. To test this, we moved in a 1-Mb sliding window along the genome (step size 100 kb) and calculated the divergence between MA lines as expected from our model after 31 generations of independent selfing (Fig. 2*B* and *SI Appendix*). Our calculations predicted that CG-methylation divergence is low in TE-rich pericentromeric regions and high in gene-rich chromosome arms (Fig. 2*B* and *SI Appendix*, Figs. S5 and S6). Remarkably, these predictions strongly agreed with the observed divergence patterns at the genome-wide scale ($R^2 = 0.74$, $P < 0.0001$).

An alternative, or complementary, explanation is that the annotation-specific divergence patterns are simply a reflection of the genome-wide distribution of heterochromatic domains, which would explain the clear partitioning between pericentromeres and euchromatin. To test this directly, we reanalyzed recent ChIP-seq data on histone variant H2A.W (21), a proxy for heterochromatin, and estimated epimutation rates for CGs in regions that were either enriched or depleted for H2A.W (*SI Appendix*). We used these rates in combination with the genome-wide density distribution of H2A.W to derive predictions of CG-methylation divergence patterns. Our analysis revealed that, at the genome-wide scale, heterochromatin-based predictions were approximately equivalent to annotation-based predictions ($R^2 = 0.72$, $P < 0.0001$, *SI Appendix*, Fig. S5), suggesting that chromatin environment is a sufficient and parsimonious explanation for the observed divergence patterns along chromosomes. These results further indicate that the maintenance of methylation at CG dinucleotides is slightly more error-prone in regions of open chromatin compared with more compact regions, probably as a by-product of active transcription.

The Spectrum of Neutral Epimutations Shapes CG Methylation Diversity in Natural Populations. An intriguing question is to what extent the epimutation landscape in the MA lines provides insights into the mechanisms that shape CG methylation diversity in *A. thaliana* natural populations, which are the outcome of long and complex evolutionary processes. To assess this we reanalyzed MethylC-seq data from a large number of accessions collected from across the Northern Hemisphere (22) (*SI Appendix*, Table S8). We focused on a subset of 133 accessions that met our quality criteria and calculated CG-methylation diversity in a 1-Mb sliding window using the same protocol as with the MA lines (*SI Appendix*). Although the natural accessions were clearly more diverse (Fig. 2*B*), genome-wide diversity patterns were highly similar to those seen in the MA lines (weighted $R^2 = 0.624$, $P < 0.0001$, Fig. 2*C* and *SI Appendix*, Fig. S7), particularly in pericentromeric regions ($R^2 = 0.899$, $P < 0.0001$) and to a slightly lesser extent in chromosome arms ($R^2 = 0.525$, $P < 0.0001$). These observations are consistent with a recent report by Hagmann et al. (23). Moreover, CG-methylation divergence among the MA lines was also moderately correlated with sequence diversity in the accessions, explaining over 25% of the genome-wide SNP distribution (weighted $R^2 = 0.254$, $P < 0.0001$, Fig. 2*D* and *SI Appendix*, Fig. S8).

It is unlikely that global patterns of CG-methylation diversity among natural accessions are the result of selection acting over broad genomic regions, because the same patterns are quickly established in isogenic MA lines in the course of only 31 generations under constant environmental conditions. Rather, our results suggest that these patterns reflect major structural properties of the *A. thaliana* genome, which modulate the ratio of forward–backward epimutation rates, and thus determine the accumulation dynamics of neutral epimutations over time. It is therefore not surprising that the reorganization of genomes during macroevolution is necessarily accompanied by a repatterning of methylation divergence among lineages or species (24), insofar that such structural changes alter genome-wide annotation densities and their accompanying chromatin environment. However, structural changes of this type are less prevalent in the course of microevolution; hence, neutral epimutations are probably the single most important factor in shaping methylome diversity in populations over short to intermediate evolutionary timescales.

Discussion

CG Epimutation Rates Are High Enough to Rapidly Uncouple Genetic and Epigenetic Variation over Evolutionary Timescales. Our analysis shows that CG epimutations are about five orders of magnitude more frequent than genetic mutations in *A. thaliana* [$\sim 10^{-4}$ compared with $\sim 10^{-9}$ (25)] and are subject to forward–backward dynamics that are rarely observed for genetic loci. Because of these properties, it is intuitively obvious that these epimutation dynamics will lead to an uncoupling of epigenetic from genetic variation over relatively short evolutionary timescales (26). Simple deterministic models show that in a strictly selfing system without selection it would require only about 800 generations to reduce correlations between genotype and epigenotype from unity to below 0.5,

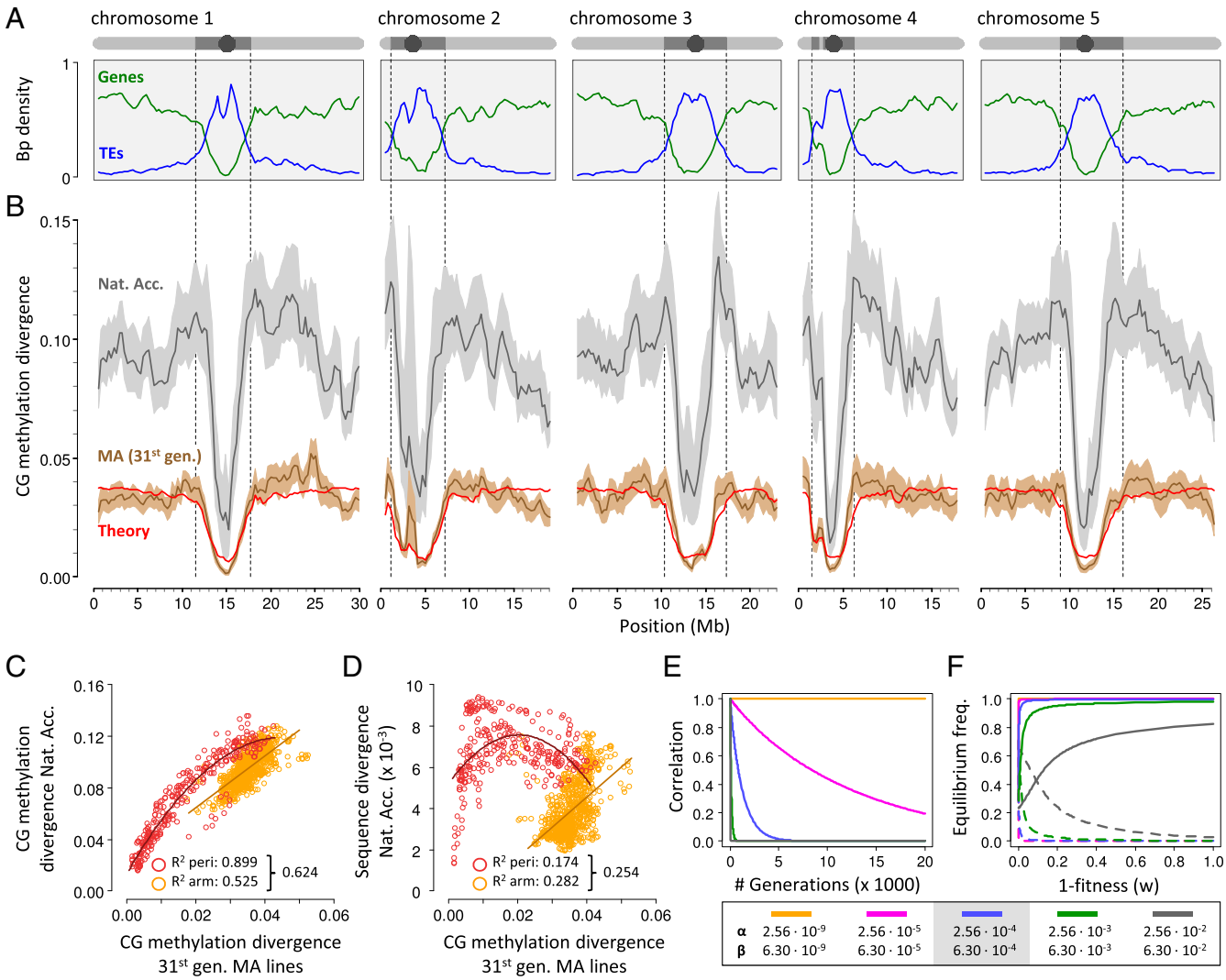

Fig. 2. (*A*) Genome-wide gene (green) and TE (blue) density as well as a schematic representation of chromosomes (circle, centromere; dark gray, pericentromeric region; light gray, arm). (*B*) Genome-wide CG methylation divergence patterns among the 31st generation MA lines (MA1_1 and MA1_2) and the natural accessions (brown and gray, respectively). The red line indicates the theoretical prediction of divergence based on the estimated epimutation rates per annotation weighted by local annotation densities. Genome-wide divergence patterns in the MA lines are strongly correlated with the diversity patterns in the natural accessions (*C*), as well as with sequence diversity in the accessions (*D*). (*E*) The relationship between genetic and epigenetic variation as a function of time and different values of the epimutation rates α and β in a strictly selfing system without selection; *x* axis, time in generations; *y* axis, expected correlation between genotype and epigenotype of two perfectly linked loci (recombination fraction = 0). The estimated CG epimutation rate (blue line) is high enough to efficiently uncouple genetic from epigenetic variation over relatively short timescales (*SI Appendix*). (*F*) Equilibrium frequency (*y* axis) of epigenotypes $c^m c^m$ (solid) and $c^u c^u$ (dashed) in a strictly selfing system as a function of fitness (*x* axis) and different forward–backward epimutation rates (colored lines). The fitness of epigenotypes $c^u c^u$, $c^u c^m$ and $c^m c^m$ was defined by *w*, $0.5(1+w)$ and 1, respectively (*SI Appendix*). The estimated CG epimutation rate (blue line) is low enough to yield epimutation-selection equilibria close to those found for DNA sequence mutation rates, even under weak selection regimes (i.e., small fitness differentials between $c^u c^u$ and $c^m c^m$). This means that CG-type epialleles should be stable enough to effectively respond to long-term selection, provided they affect fitness.

and only about 2,700 generations to reduce it to below 0.1 (Fig. 2*E* and *SI Appendix*), and this breakdown is expected to be even faster in outcrossing systems. This rapid uncoupling may explain why variation in DNA methylation in *A. thaliana* populations is only partly associated with *cis*- and *trans*-acting DNA sequence variants (22), and thus sheds new light on the molecular mechanisms that drive the coevolution of genomes and epigenomes. One situation in which genotype-epigenotype associations are expected to be more prevalent is when natural accessions have only newly diverged from a common ancestor, as it may be the case in recently founded local populations. This prediction can be tested using genome-wide association study-based *cis*- and *trans*-mapping analysis across different groups of accessions that vary along a gradient of genetic relatedness and (or) geographic locations. Recent whole-genome and methylome datasets of *A. thaliana* local populations collected in North America (23) and Sweden (27) may be applicable for that purpose.

CG Epimutation Rates Are Low Enough for New Epialleles to Sustain Long-Term Selection Responses. Although our results provide strong evidence that global patterns of CG-methylation diversity among *A. thaliana* natural accessions are mainly influenced by the accumulation of selectively neutral epimutations, targeted selection of epialleles at specific loci may still be an important process. Particularly in chromosome arms, the MA divergence patterns were only moderately correlated with those of the natural accessions, suggesting the involvement of other factors such as direct selection on CG methylation states and (or) selection via DNA sequence

variants that indirectly regulate CG methylation in *cis* or *trans*. Indeed, the observation that methylation profiles of orthologous genes is often highly conserved across species (28) indicates that some epigenetic states are subject to strong evolutionary constraints. For epigenetic selection to be effective, epimutations need to be sufficiently stable (29), and a lack of stability has been cited as one reason why epigenetic inheritance has no potent role in evolution or in the heritability of complex traits (30). Contrary to these conclusions, simple deterministic selection models show that newly arising epimutations are stable enough to respond effectively to long-term selection, even under weak selection regimes, yielding epimutation-selection equilibria that are close to those expected for DNA sequence mutation rates (Fig. 2*F* and *SI Appendix*).

Reference Values for Future Population Epigenetic Studies. In light of our estimates of forward–backward epimutation rates, future work should examine the effect of selection in more complex population genetic models that account for finite population sizes, migration, and drift such as those proposed by Charlesworth and Jain (13). Recently, Wang and Fan (14) devised a neutrality test based on single methylation polymorphism data using a modified version of Tajima's *D*. We caution that care needs to be taken when supplying epimutation rates to this or similar tests. Incorrect assumptions about the ratio of forward and backward rates can lead to widely misleading conclusions regarding the role of selection on CG methylation. If one assumes that forward and backward rates are equivalent, TE-associated CGs would most likely be detected as being under strong selection, and pericentromeric regions would seem to have undergone selective sweeps. However, if one considers that spontaneous methylation gain is about 30 times more likely than methylation loss (see Table 1), equilibrium levels of CG-methylation diversity in TEs would seem to be entirely consistent with neutrality. Hence, the context- or annotation-specific epimutation rates provided here should serve as useful reference values when inferring signatures of epigenetic selection in *A. thaliana* and possibly in other plant species.

Materials and Methods

Below we provide a brief description of the theoretical model and our estimation approach. For a more detailed explanation we refer the reader to *SI Appendix*.

Derivation of Neutral Epimutation Model. Let c^u and c^m denote an unmethylated and a methylated cytosine, respectively, and $\alpha = Pr(c^u \to c^m)$ and $\beta = Pr(c^m \to c^u)$ be the probabilities that a cytosine gains or loses methylation during or before gamete formation, which can include gains or losses of DNA methylation in somatic tissues from which the gametic cells were derived. We arbitrarily call α the forward and β the backward epimutation rate per generation per haploid methylome. We modeled the epigenotype frequencies at the *j*th cytosine using a Markov chain with three states: c^uc^u, c^uc^m, and c^mc^m. Taking into account Mendelian segregation of epialleles c^m and c^u together with rates α and β, we derived the epigenotype transition matrix **T** after one selfing generation:

	c^uc^u	c^mc^u	c^mc^m
c^uc^u	$(1-\alpha)^2$	$2(1-\alpha)\alpha$	α^2
c^uc^m	$\frac{1}{4}(\beta+1-\alpha)^2$	$\frac{1}{2}(\beta+1-\alpha)(\alpha+1-\beta)$	$\frac{1}{4}(\alpha+1-\beta)^2$
c^mc^m	β^2	$2(1-\beta)\beta$	$(1-\beta)^2$

This formulation does not account for higher-order epimutation events, because such events are expected to be rare for small epimutation rates. Following Markov chain theory, the epigenotype frequencies at cytosine *j* in the MA population after *t* generations of single seed descent, π_{tj}, can be expressed as $\pi_{tj} = \pi_{0j}\mathbf{P}\mathbf{V}^t\mathbf{P}^{-1}$, where *P* is the eigenvector of matrix **T** and **V** is a diagonal matrix of the eigenvalues of matrix **T**. Using Mathematica 10.0 (Wolfram Research, Inc.) we derived analytical solutions for the elements of π_{tj}, which are functions of *t*, α, β as well as the initial frequency vector π_{0j}. These analytical solutions have no easy form and are therefore omitted here for brevity. At equilibrium, the $\pi_{\infty j}$ represent the expected epigenotype frequencies at cytosine *j* among the MA lines after a (hypothetical) infinite number of selfing generations $(t=\infty)$, and were obtained by calculating $\lim_{t\to\infty}\pi_{tj}$:

$$\pi_{\infty j}(c^uc^u) = \frac{\beta\left((1-\beta)^2-(1-\alpha)^2-1\right)}{(\alpha+\beta)\left((\alpha+\beta-1)^2-2\right)}$$

$$\pi_{\infty j}(c^uc^m) = \frac{4\alpha\beta(\alpha+\beta-2)}{(\alpha+\beta)\left((\alpha+\beta-1)^2-2\right)}$$

$$\pi_{\infty j}(c^mc^m) = \frac{\alpha\left((1-\alpha)^2-(1-\beta)^2-1\right)}{(\alpha+\beta)\left((\alpha+\beta-1)^2-2\right)}.$$

For any $0<\alpha,\beta<1$, these equilibrium solutions are independent of the initial epigenotype proportions π_{0j} in the common founder, and depend only on the rates α and β. The rate at which the epigenotype proportions converge to these equilibrium values depends on the relative magnitude of the forward and backward rates.

Modeling Methylation Divergence. To derive analytical formulas for methylation divergence, we score the methylation divergence between two independently selfed lines at every cytosine with the following distance matrix:

	c^uc^u	c^mc^u	c^mc^m
c^uc^u	0	$\frac{1}{2}$	1
c^uc^m	$\frac{1}{2}$	0	$\frac{1}{2}$
c^mc^m	1	$\frac{1}{2}$	0

Let t_1 and t_2 denote the number of generations between two individuals at generations G_m and G_n and their most recent common founder at generation G_f, respectively (i.e., $t_1 = G_m - G_f$, $t_2 = G_n - G_f$, Fig. 1*A*). Let $\pi_{t_ij}|c^mc^m$ be the vector of epigenotype frequencies at the *j*th cytosine after t_i selfing generations from G_f, conditional on the fact that the most recent common founder epigenotype was c^mc^m: $\pi_{t_ij}|c^mc^m = (0,0,1)\cdot T^{t_i}$. Using this equation and the methylation divergence scoring table above, the divergence between these two lines at this locus can be calculated as

$$d_{t_1t_2j}|c^mc^m = \frac{1}{2}\sum_{k=1}^{4}\left(\pi_{t_1j}(P1_k)|c^mc^m\cdot\pi_{t_2j}(P2_k)|c^mc^m\right) + 1\sum_{k=1}^{2}\left(\pi_{t_1j}(Q1_k)|c^mc^m\cdot\pi_{t_2j}(Q2_k)|c^mc^m\right),$$

with $Q1=\{c^uc^u, c^mc^m\}$, $Q2=\{c^mc^m, c^uc^u\}$, $P1=\{c^uc^u, c^uc^m, c^uc^m, c^mc^m\}$, and $P2=\{c^uc^m, c^uc^u, c^mc^m, c^uc^m\}$. The simple multiplication of these frequencies follows from the fact that the selfing lines are conditionally independent. The divergence over all loci for which the most recent common founder at G_f was c^mc^m is

$$d_{G_f,t_1t_2}|c^mc^m = \sum_j d_{t_1t_2j}|c^mc^m = N^{mm}_{G_f}\cdot d_{t_1t_2j}|c^mc^m,$$

where $N^{mm}_{G_f}$ are the number of methylated cytosines at G_f. The global (or total) DNA methylation divergence along the genome can be calculated as

$$D_{G_f,t_1t_2} = d_{G_f,t_1t_2}|c^mc^m + d_{G_f,t_1t_2}|c^uc^m + d_{G_f,t_1t_2}|c^uc^u,$$

where $d_{G_f,t_1t_2}|c^uc^m$ and $d_{G_f,t_1t_2j}|c^uc^u$ are derived using similar arguments as for $d_{G_f,t_1t_2}|c^mc^m$. We prefer to express the global methylation divergence as a proportion of all of the cytosines, in which case

$$D^*_{G_f,t_1t_2} = \frac{D_{G_f,t_1t_2}}{N}.$$

Using the above derived equilibrium epigenotype frequencies, it can be shown that the equilibrium divergence is

$$D^*_{\infty}=\frac{2\alpha\beta\left(\left[(1-\beta)^2-(1-\alpha)^2\right]^2-2[\alpha+\beta-1]^2+3\right)}{(\alpha+\beta)^2\left((\alpha+\beta-1)^2-2\right)^2}.$$

Model Fitting and Parameter Estimation. For each pedigree we had a number M of line comparisons and we denoted the observed methylation divergence between each of them as O_{G_f,t_1t_2i}, with $i=\{1,2,\ldots,M\}$, and G_f, t_1, and t_2 the times of and from their most recent common founder, respectively. We assumed that these observations were generated from the proposed epimutation model but contained some unknown measurement error. Hence, we had

$$O_{G_f,t_1t_2i}=c+D^*_{G_f,t_1t_2}+\epsilon_i,$$

where c is the intercept, $D^*_{G_f,t_1t_2}$ is the theoretical global divergence measure introduced above, and ϵ is a random measurement error term. For the MA1_1 population the value of c was approximated using the methylation divergence between technical replicates. For the other three populations no technical replicates were available and c was estimated along with the other parameters (*SI Appendix*, Fig. S9). To obtain parameter estimates we minimized $r^2=\sum_i\left(O_{G_f,t_1t_2i}-D^*_{G_f,t_1t_2}\right)^2$, which is a problem in multivariate nonlinear regression. This involves finding solutions to $\nabla r^2=0$, which can be obtained numerically. Extensive simulations showed that our estimation method performs well, even with relatively large measurement error (*SI Appendix*, Fig. S10).

ACKNOWLEDGMENTS. We thank B. Charlesworth and J. Hadfield for their comments during a seminar at the University of Edinburgh. This work was supported by grants from the Netherlands Organization for Scientific Research (to R.C.J., R.W., A.v.d.G., F.J., and M.C.-T.), a University of Groningen Rosalind Franklin Fellowship (to M.C.-T.), National Institutes of Health Grant R00GM100000, and National Science Foundation Grant IOS-1339194 (to R.J.S.).

1. Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11(3):204–220.
2. Becker C, et al. (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature* 480(7376):245–249.
3. Schmitz RJ, et al. (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science* 334(6054):369–373.
4. Richards EJ (2006) Inherited epigenetic variation—revisiting soft inheritance. *Nat Rev Genet* 7(5):395–401.
5. Cortijo S, et al. (2014) Mapping the epigenetic basis of complex traits. *Science* 343(6175):1145–1148.
6. Kalisz S, Purugganan MD (2004) Epialleles via DNA methylation: Consequences for plant evolution. *Trends Ecol Evol* 19(6):309–314.
7. Weigel D, Colot V (2012) Epialleles in plant evolution. *Genome Biol* 13(10):249.
8. Diez CM, Roessler K, Gaut BS (2014) Epigenetics and plant genome evolution. *Curr Opin Plant Biol* 18:1–8.
9. Springer NM (2013) Epigenetics and crop improvement. *Trends Genet* 29(4):241–247.
10. Jiang C, Mithani A, Belfield EJ, Mott R, Hurst LD, Harberd NP (2014) Environmentally responsive genome-wide accumulation of de novo *Arabidopsis thaliana* mutations and epimutations. *Genome Res* 24(11):1821–1829.
11. Becker C, Weigel D (2012) Epigenetic variation: Origin and transgenerational inheritance. *Curr Opin Plant Biol* 15(5):562–567.
12. Hunter B, Hollister JD, Bomblies K (2012) Epigenetic inheritance: What news for evolution? *Curr Biol* 22(2):R54–R56.
13. Charlesworth B, Jain K (2014) Purifying selection, drift, and reversible mutation with arbitrarily high mutation rates. *Genetics* 198(4):1587–1602.
14. Wang J, Fan C (2015) A neutrality test for detecting selection on DNA methylation using single methylation polymorphism frequency spectrum. *Genome Biol Evol* 7(1):154–171.
15. Shaw RG, Byers DL, Darmo E (2000) Spontaneous mutational effects on reproductive traits of *Arabidopsis thaliana*. *Genetics* 155(1):369–378.
16. Lister R, et al. (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133(3):523–536.
17. Cokus SJ, et al. (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452(7184):215–219.
18. Nuthikattu S, et al. (2013) The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotide small interfering RNAs. *Plant Physiol* 162(1):116–131.
19. Zemach A, et al. (2013) The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153(1):193–205.
20. Creasey KM, et al. (2014) miRNAs trigger widespread epigenetically activated siRNAs from transposons in *Arabidopsis*. *Nature* 508(7496):411–415.
21. Yelagandula R, et al. (2014) The histone variant H2A.W defines heterochromatin and promotes chromatin condensation in *Arabidopsis*. *Cell* 158(1):98–109.
22. Schmitz RJ, et al. (2013) Patterns of population epigenomic diversity. *Nature* 495(7440):193–198.
23. Hagmann J, et al. (2015) Century-scale methylome stability in a recently diverged *Arabidopsis thaliana* lineage. *PLoS Genet* 11(1):e1004920.
24. Seymour DK, Koenig D, Hagmann J, Becker C, Weigel D (2014) Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization. *PLoS Genet* 10(11):e1004785.
25. Ossowski S, et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327(5961):92–94.
26. Johannes F, Colot V, Jansen RC (2008) Epigenome dynamics: A quantitative genetics perspective. *Nat Rev Genet* 9(11):883–890.
27. Dubin MJ, et al. (2015) DNA methylation variation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. arXiv:1410.5723.
28. Takuno S, Gaut BS (2013) Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proc Natl Acad Sci USA* 110(5):1797–1802.
29. Furrow RE (2014) Epigenetic inheritance, epimutation, and the response to selection. *PLoS ONE* 9(7):e101559.
30. Slatkin M (2009) Epigenetic inheritance and the missing heritability problem. *Genetics* 182(3):845–850.