Genetic and environmental drivers of large-scale epigenetic variation in *Thlaspi arvense*

Dario Galanti[1], Daniela Ramos-Cruz[2,3], Adam Nunn[4,5], Isaac Rodríguez-Arévalo[2,3], J.F. Scheepens[6], Claude Becker[2,3], Oliver Bossdorf[1]

[1] Plant Evolutionary Ecology, Institute of Evolution and Ecology, University of Tübingen, 72076 Tübingen, Germany.

[2] Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, Vienna BioCenter (VBC), 1030 Vienna, Austria.

[3] Genetics, Faculty of Biology, Ludwig Maximilians University Munich, 82152 Martinsried, Germany.

[4] ecSeq Bioinformatics GmbH, Leipzig 04103, Germany.

[5] Institute for Computer Science, University of Leipzig, Leipzig 04107, Germany.

[6] Plant Evolutionary Ecology, Institute for Ecology, Evolution and Diversity, Faculty of Biological Sciences, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany.

Keywords: DMRs, DNA methylation, field pennycress, GWAS, landscape genomics, population epigenetics, WGBS.

Abstract

Natural plant populations often harbour substantial heritable variation in DNA methylation. However, a thorough understanding of the genetic and environmental drivers of this epigenetic variation requires large-scale and high-resolution data, which currently exist only for a few model species. Here, we studied 207 lines of the annual weed *Thlaspi arvense* (field pennycress), collected across a large latitudinal gradient in Europe and propagated in a common environment. By screening for variation in DNA sequence and DNA methylation using whole-genome (bisulfite) sequencing, we found significant epigenetic population structure across Europe. Average levels of DNA methylation were strongly context-dependent, with highest DNA methylation in CG context, particularly in transposable elements and in intergenic regions. Residual DNA methylation variation within all contexts was associated with genetic variants, which often co-localized with annotated methylation machinery genes but also with new candidates. Variation in DNA methylation was also significantly associated with climate of origin, with methylation levels being higher in warmer regions and lower in more variable climates. Finally, we used variance decomposition to assess genetic versus environmental associations with differentially methylation regions (DMRs). We found that while genetic variation was generally the strongest predictor of DMRs, the strength of environmental associations increased from CG to CHG and CHH, with climate-of-origin as the strongest predictor in about one third of the CHH DMRs. In summary, our data show that natural epigenetic variation in *Thlaspi arvense* is significantly associated with both DNA sequence and environment of origin, and that the relative importance of the two factors strongly depends on the sequence context of DNA methylation. *T. arvense* is an emerging biofuel and winter cover crop; our results may hence be relevant for breeding efforts and agricultural practices in the context of rapidly changing environmental conditions.

Author Summary

Variation within species is an important level of biodiversity, and it is key for future adaptation. Besides variation in DNA sequence, plants also harbour heritable variation in DNA methylation, and we want to understand the evolutionary significance of this epigenetic variation, in particular how much of it is under genetic control, and how much is associated with the environment. We addressed these questions in a high-resolution molecular analysis of 207 lines of the common plant field pennycress (*Thlaspi arvense*), which we collected across Europe, propagated under standardized conditions, and sequenced for their genetic and epigenetic variation. We found large geographic variation in DNA methylation, associated with both DNA sequence and climate of origin. Genetic variation was generally the stronger predictor of DNA methylation variation, but the strength of environmental association varied between different sequence contexts. Climate-of-origin was the strongest predictor in about one third of the differentially methylated regions in the CHH context, which suggests that epigenetic variation may play a role in the short-term climate adaptation of pennycress. As pennycress is currently being domesticated as a new biofuel and winter cover crop, our results may be relevant also for agriculture, particularly in changing environments.

Introduction

Besides variation in DNA sequence, natural plant populations usually also harbour variation in epigenetic modifications of the DNA. This is particularly well documented for DNA methylation, usually referring to the addition of a methyl group to the 5th atom of the cytosine ring, a modification associated with silencing of transposable elements (TEs) and the regulation of gene expression. Variation in DNA methylation can arise if methylation marks are altered by chance during mitosis or meiosis (epimutations) (1,2), or if they are induced in response to environmental changes (3,4). Some DNA methylation differences are stably inherited through meiosis, which has led some to hypothesize that DNA methylation variation could be under natural selection and contribute to adaptation (5–7). These ideas are fuelled by the observation that DNA methylation variation in natural plant populations is often non-random and geographically structured (8–12). However, the DNA methylation variation observed in the field is always a combination of stable (= heritable) and plastic (= non-heritable) components. In order to tease these apart and describe the heritable component of DNA methylation variation, one must analyse the offspring of different populations grown in a common environment. To date, common-environment analyses of natural DNA methylation variation that cover many populations and broad environmental gradients are still rare.

In plants, DNA methylation can occur in the three sequence contexts: CG, CHG and CHH (where H is A, T or C). Distinguishing between these contexts is sensible because they differ in the molecular machineries for depositing, maintaining and removing methylation (13,14), which has consequences for their dynamics and stability. In *Arabidopsis thaliana*, CG methylation (mCG) is mostly maintained in a copy-paste manner during replication, CHG methylation (mCHG) by DNA-histone methylation self-reinforcing loops and CHH methylation (mCHH) by recursive *de-novo* methylation deposited by the RNA-directed DNA methylation pathway (RdDM) (13,14). In addition, CHG and CHH methylation partially share maintenance pathways (15,16). Overall, there is a gradient of similarity and decreasing stability from CG to CHG to CHH. Although less stable, CHH is the most abundant context and often the most responsive to stresses (17). Besides the sequence contexts, the dynamics of DNA methylation also strongly depend on the genomic features in which it occurs. While heterochromatic regions and TEs are usually heavily methylated to repress transcription, methylation is often lower and more variable in genes and regulatory regions (18–20). In addition, while DNA methylation is almost exclusively a repressive mark on TEs and in regulatory regions, this is not as clear for gene body methylation (GBM), as several constitutively expressed housekeeping genes often harbour CG but not CHG and CHH methylation (20,21). If methylation in different genomic features has different functions, then also different selective pressures are to be expected (22). Finally, for both influences

of sequence context and genomic features on methylation variation, there appears to be high species-specificity in plants (20).

To study such complex dynamics, DNA methylation can be quantified at different levels, from global (or genome-wide) methylation, to average methylation limited to sequence contexts or genomic features, to the methylation of genomic regions or individual positions. While genetic single nucleotide polymorphisms (SNPs) can have large effects, this does not seem to be the case for DNA methylation polymorphisms, which affect transcription only when accumulating over a broader genomic region (23–25). For this reason, the study of differentially methylated regions (DMRs) became very popular in high-resolution studies (11,12,18,26).

Given the complex molecular machinery for regulation and maintenance of DNA methylation, it is not surprising that previous studies have demonstrated various kinds of genetic control over DNA methylation variation. Genetic polymorphisms can control DNA methylation in *cis,* for example, when a TE insertion next to a gene promoter induces the methylation of the latter (25), or in *trans*, when genetic mutations affect genes involved in the DNA methylation machinery (11,12,27). In the latter case, variation in individual DNA loci often affects methylation levels across the entire genome. In addition, a number of genes have been found to affect methylation levels indirectly, acting upstream or in aid of the methylation machinery. In particular, ubiquitination, a post-translational modification affecting histone tails and protein turnover, affects DNA methylation in plants and animals in several ways (28–33). For example, in plants ORTH/VIM E3 ubiquitin ligases recruit DMT1 for methylation maintenance through ubiquitination of histone tails (30,31). However, in spite of this functional understanding of several mechanisms of genetic control, we still lack a good understanding of the degree of genetic determination of DNA methylation variation in wild plant populations.

If DNA methylation variation is under natural selection – whether independently from DNA sequence or linked to it – we expect this to result in patterns of association between methylation variation and the environment. Several previous studies indeed found correlations between methylation patterns and habitat or climate in different plant species (8–12,34). However, most of these studies were either conducted in the field, based on only few natural populations, or used low-resolution molecular methods, which limited their generalizability and/or their power to detect environment-methylation associations and to separate genetically controlled from independent components of DNA methylation variation (5). The only available data that does not suffer from any of these limitations comes from *Arabidopsis thaliana* (11,12,18), a plant with an exceptionally small and simple genome, with low numbers of TEs, and low global DNA methylation (35). To advance our understanding of natural DNA methylation variation and its genetic and environmental drivers, we must expand our scope and collect large-scale, high-resolution data also for other plant species.

Here, we present a detailed genomic analysis of 207 lines of the plant *Thlaspi arvense* (field pennycress) that we collected across a latitudinal gradient in Europe, cultivated in a common environment, and profiled for genomic and epigenomic variation. Like *A. thaliana*, *T. arvense* is an annual and mostly selfing member of the Brassicaceae family, but it has a significantly larger genome of approx. 500Gb, which is richer in TEs and DNA methylation (36). The species is an interesting study object also because it is currently being domesticated into a new biofuel and cover crop (37–41). The genomic work with *T. arvense* is facilitated by recently published high-quality reference genomes (36,42). In our study, we demonstrate that European populations of *T. arvense* harbour substantial natural epigenetic variation, which is associated with DNA sequence variation as well as with climate of origin, but in a highly context-dependent manner. In our data, genetic variation was generally the stronger predictor of DNA methylation variation. Genome-wide association analyses identified several candidate loci, but there was a fraction of the DNA methylation variation that was most strongly associated with climate of origin, suggesting a link with climate adaptation.

Results

The 207 *Thlaspi arvense* lines we worked with came from 36 natural populations which we sampled across Europe in 2018, on a latitudinal gradient from Southern France to Central Sweden, with three populations each in Southern France and The Netherlands, seven in Southern Germany, eight in Central Germany and South Sweden, respectively, and another seven populations in Central Sweden (Fig 1A and Table S1). In each population, we collected seeds of 4-6 different lines (Table S1). We grew all lines under common environmental conditions, extracted their DNA and generated Whole Genome Sequencing (WGS) and Whole Genome Bisulfite Sequencing (WGBS) libraries, which were sequenced with an average coverage of 25.3x and 30.8x, respectively. Variant calling retrieved around nine million SNPs and short INDELs with genotypes called in >90% of the lines. The methylation calling retrieved about 16 million, 18.4 million and 95.3 million positions in CG, CHG and CHH contexts, respectively, with up to 25% missing values. The global DNA methylation, calculated by averaging the ratio between methylated and total reads at every analysed cytosine, was estimated at 18.5% (average of all lines).

We found significant genetic and epigenetic population structure across Europe. A principal component analysis (PCA) based on genetic variants showed two main clades: a larger one including almost all lines from France, Germany and the Netherlands, and a smaller one that consisted almost exclusively of Swedish lines (Fig 1B). The larger clade also showed a clear latitudinal gradient. PCAs based on DNA methylation variation generally also found two major clades, with the CG methylation-based patterns most closely resembling the genetically-based ones, and a decreasing similarity

between genetic and epigenetic population structure from CG to CHG to CHH methylation (Fig 1B and S1). Restricting methylation to specific genomic features also revealed that mCG of genes and promoters has stronger geographic patterns that methylation of TEs (Fig S1).

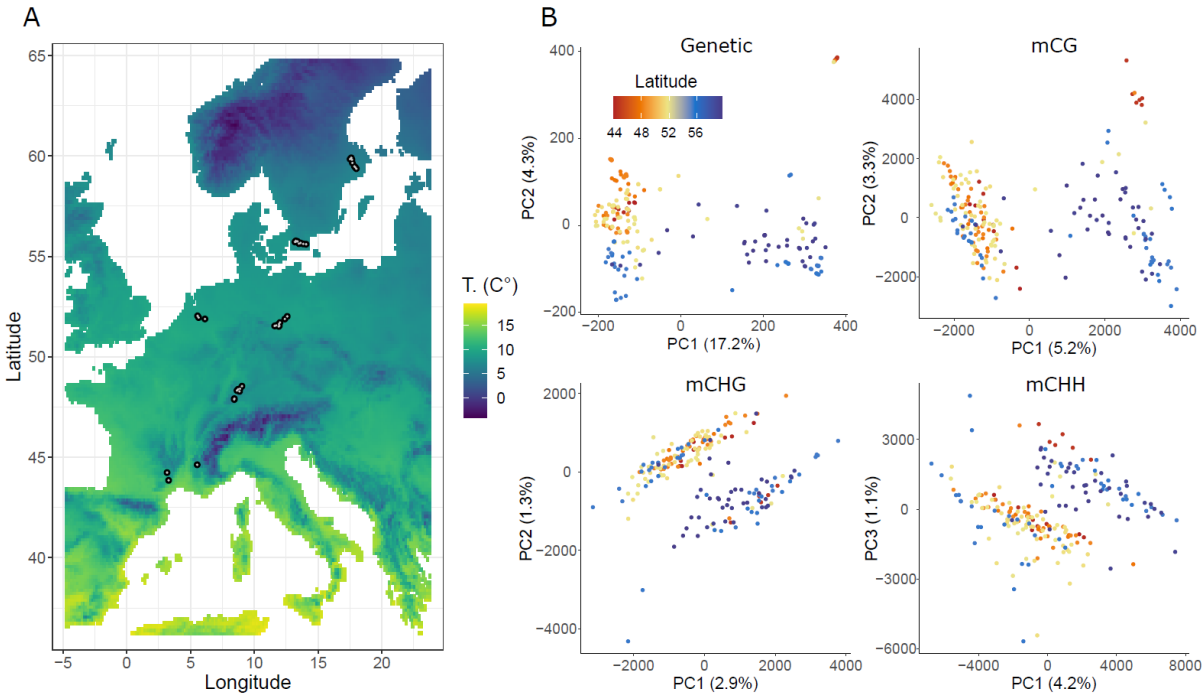


Fig 1. Geographic distribution and population structure of the 207 sampled *Thlaspi arvense* lines. (A) Geographic locations of the 36 populations, with background colours indicating mean annual temperature (T.). (B) PCA plots of all 207 lines based on DNA sequence ("Genetic") and DNA methylation in different sequence contexts ("mCG", "mCHG" and "mCHH").

Average methylation

To understand the structure of DNA methylation variation in *T. arvense*, we first examined patterns of average methylation across all lines. We not only distinguished between the three sequence contexts CG, CHG and CHH, but we also assigned cytosines to different genomic features: genes, promoters, TEs and intergenic regions. For genes and TEs, we used available annotations (36), while for promoters we considered the 2 kb upstream sequences of genes (or until the boundary of the previous gene if closer). We considered intergenic space, anything not belonging to these categories. Across all genomic features, the average methylation was much higher in CG context than in CHG and CHH; for the latter two it was generally similar (Fig 2A). TEs were the most highly methylated genomic features, followed by intergenic regions, whereas promoters and especially gene bodies showed very low average methylation (Fig 2A). For instance, while for CG sites in TEs the average methylation was around 80%, it was below 2% for CHH sites in genes. Although these patterns are conserved in the whole collection, there is large residual variation between lines, which is particularly high in TEs (up to 12%) and decreases gradually moving to intergenic regions, promoters and particularly genes (Fig 2A). Finally, partially due to TEs covering about 60% of the *T. arvense* genome

(36), its global methylation of 18.5% (average of all lines) is much higher than that of *A. thaliana* (5.8%)(12) and many other Brassicaceae (20).

To better understand the observed values of average methylation, and in particular the low gene body methylation, we further examined the distributions of methylation values of individual genes, TEs and promoters, averaging across all lines. Interestingly, while context-specific methylation levels were very consistent for TEs, almost exclusively methylated, we found bimodal distributions for genes and promoters, with a large majority of unmethylated and a smaller fraction of methylated features (Fig 2B). Intersecting genes methylated in each context (using mCG > 20% and mCHG/mCHH > 8% as cutoffs) we confirmed that a large portion of these is methylated in all context, showing a TE-like methylation signature, and a much smaller portion only in CG (Fig S2A)(36). Even though many TE-like methylated genes might be pseudogenes, a gene ontology (GO) enrichment analysis surprisingly found enrichment for a variety of housekeeping-like GO terms such as photosynthesis, protein modification and nucleic acid processing (Fig S2B). Instead, the small portion of genes methylated only in CG, were enriched for pathogen defence functions (Fig S2B).

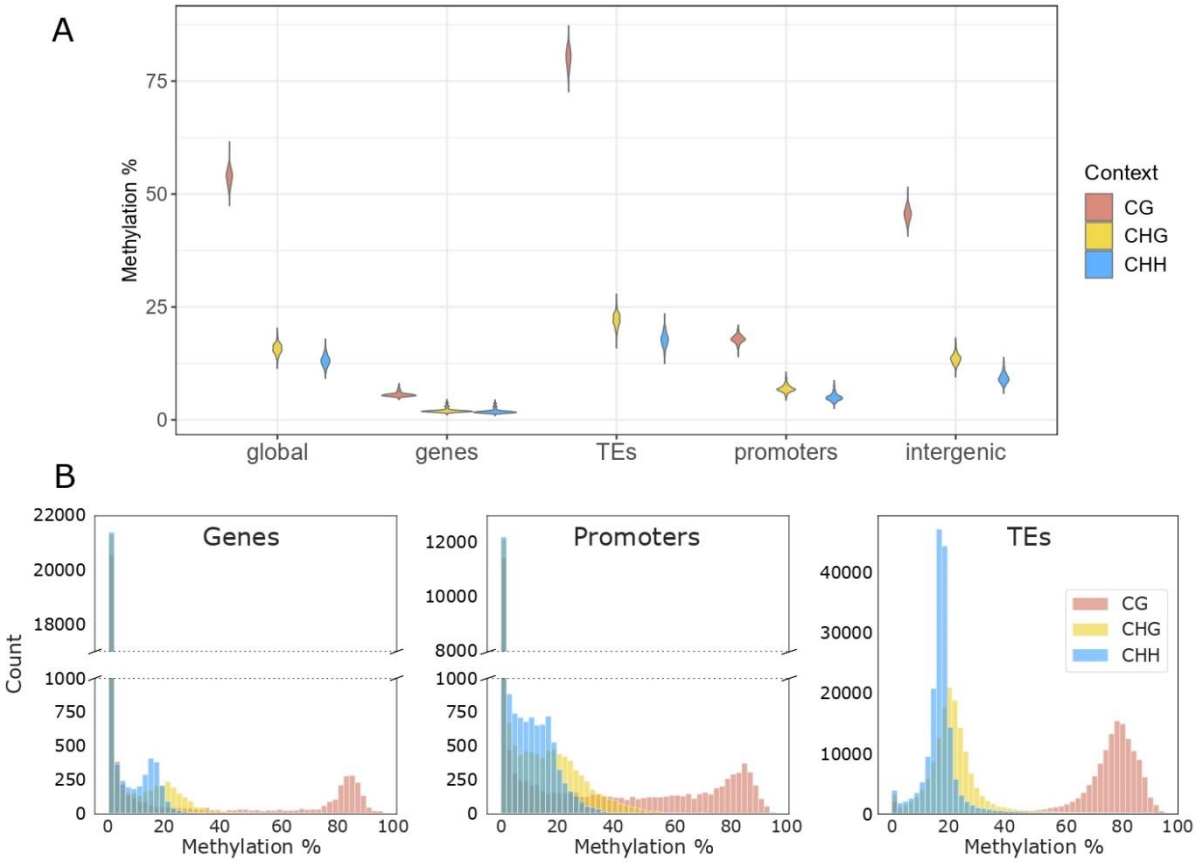


Fig 2. Average methylation and distributions of methylation values for different sequence contexts and genomic features in *T. arvense*. (A) Average methylation levels of genomic features; violin plots represent variation between lines. (B) Distributions of individual methylation values for genes, promoters and transposable elements (TEs) obtained averaging across all 207 lines.

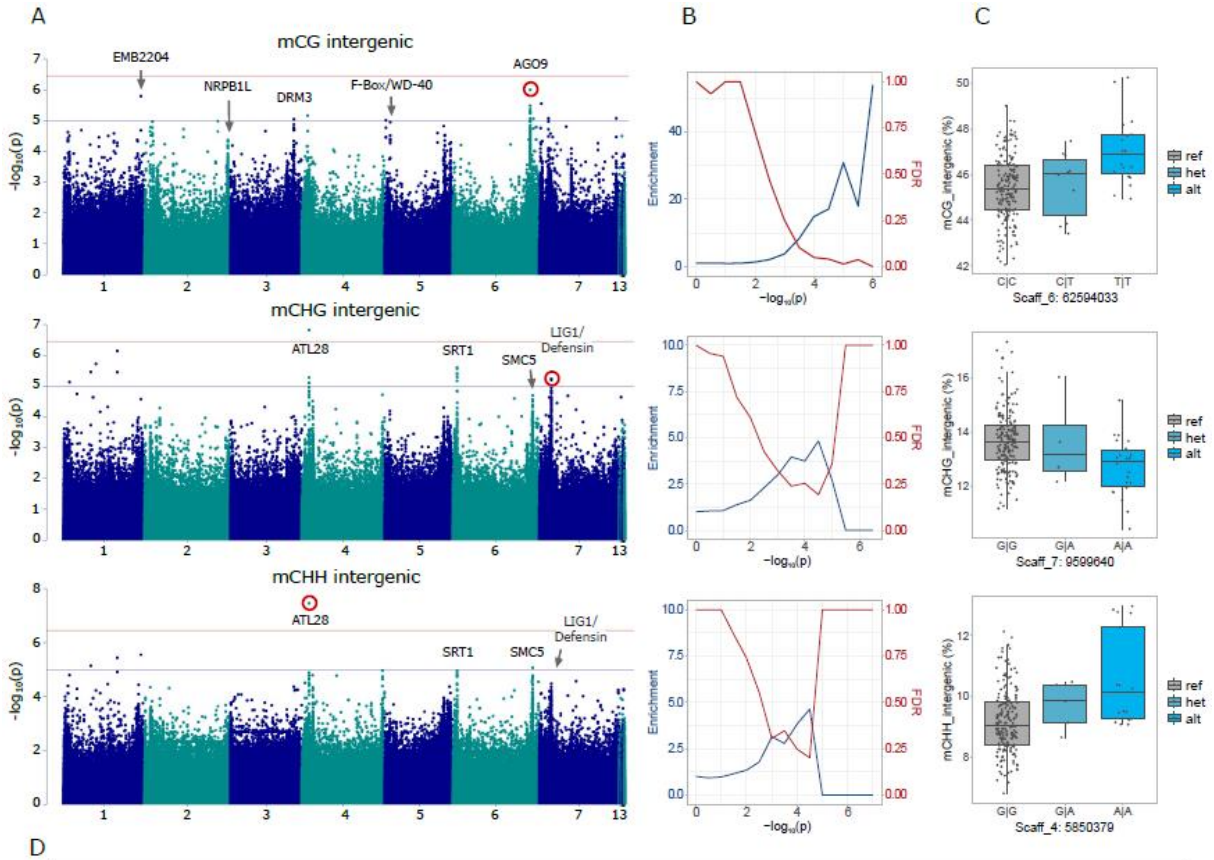
Genetic basis of methylation variation

To understand the genetic basis of the observed methylation variation, we employed genome-wide association (GWA) analyses that tested for statistical associations between every biallelic genetic variant and the average methylation of every sequence context and genomic feature. We restricted our analyses to genetic variants with a minor allele frequency (MAF) ≥ 0.04; however, repeating all analyses with a MAF>0.01 did not influence the results relevantly. Since large numbers of unmethylated genes (Fig 2B) could potentially obscure association patterns in methylated genes, we re-ran these analyses for average methylation levels based only on genes with a GBM > 5%. In all GWA analyses, we corrected for population structure using an Isolation-By-State (IBS) distance matrix. Although our experimental design and number of lines hardly provided sufficient power to meet a full Bonferroni threshold, we found that many of the genetic variants that were most strongly associated with methylation levels were close to genes with predicted functions related to DNA methylation (Figs. 3A, 3D and S3). For instance, one strong candidate was an orthologue of *ARGONAUTE 9 (AGO9)*, coding a DICER-like protein involved in RNA silencing; *AGO9* natural variation is associated with mCHH in TEs in *A. thaliana* (12). Another candidate was an orthologue of *DOMAINS REARRANGED METHYLTRANSFERASE 3 (DRM3)*, which is involved in RdDM and non-CG methylation maintenance in Arabidopsis (43–45). Reflecting the multigenic basis of methylation, even the higher -log(p) variants had relatively small size effects of about 1.5% methylation (Fig 3C).

To confirm the suspected enrichment of methylation-related genes among stronger associations, we conducted an enrichment analysis based on all genetic variants within 20kb from *a priori* candidate genes – orthologues of *A. thaliana* genes known to affect methylation (Table S3). For some genomic features and sequence contexts, we indeed found an enrichment of these *a priori* candidates among the genetic variants most strongly associated with average methylation levels (e.g. mCG in Fig 3B), but in many cases the top variants were not neighbouring any *a priori* candidates (drop of the enrichment for high -log(p) thresholds in mCHG and mCHH in Fig 3B; see Fig S3 for more results). Nevertheless, a search of the neighbouring regions of these variants identified several new candidates that may not affect methylation directly, but have predicted functions with a potential for indirect effects on DNA methylation. These include e.g. the histone deacetylase *SIRTUIN 1 (SRT1)*, the DNA-damage-repair/toleration *(DRT111)* and several E3 ubiquitin ligases such as F-box transcription factors and RING-H2 finger proteins (Fig 3; see Table S4 for all genes located within 15kb from variants significant at -log(p) > 5). Overall, our results showed that natural DNA methylation variation in *T. arvense* was significantly associated with underlying DNA sequence variation, but only some of the top genetic variants were known methylation machinery genes, whereas there were many

additional, less well-characterized genes that appeared to play a role, possibly through indirect effects on methylation.

The GWA results strongly differed between sequence contexts, with a unique profile of genetic variants associated with average mCG, while the results were very similar for mCHG and mCHH (Fig 3A, 3D and S3). In mCG, some of the top candidates were *AGO9*, the methyltransferase *DRM3*, the RdDM polymerase subunit *NRPB10L* and few ubiquitination related genes such as *ARABIDOPSIS TOXICOS EN LEVADURA* (*ATL) 28* and *50*. In mCHG and mCHH, the strongest associations included *SRT1*, the DNA-repair related *STRUCTURAL MAINTENANCE OF CHROMOSOMES 5 (SMC5)*, the *DNA LIGASE 1 (LIG1)*, involved in DNA demethylation, and in particular *ATL28*, which is neighbouring the only variant reaching genome-wide significance for several average methylation phenotypes.



D

Candidate gene	Gene ID	Ortholog	Putative function	Top SNP location
F-Box/WD-40	Ta_02099	AT3G52030	Histone H3-K4 methylation	<1kb upstream
AGO9	**Ta_05665**	**AT5G21150**	**Argonaute protein (siRNA silencing)**	**14kb downstream**
DRM3	**Ta_08517**	**AT3G17310**	**Methyltransferase (RdDM)**	**13kb upstream**
NRPB1L	Ta_12031	AT4G35800	Subunit of Pol II, Pol IV and Pol V (RdDM)	2kb upstream
EMB2204	Ta_21823/ 21824	AT1G22090	EMBRYO DEFECTIVE 2204	6kb upstream
SRT1	Ta_00697	AT5G55760	Sirtuin family HDAC (histone deacetylase)	1kb downstream
LIG1/ Defensin	**Ta_17381**/ Ta_17386	**AT1G08130**/ AT4G18823	**DNA Ligase1 (DNA demethylation)**/ Defensin-like protein	**15kb downstream of LIG1** / <1kb upstream of Defensin-like
SMC5	Ta_06084	AT5G15920	Structural Maintenance Of Chromosomes (DNA repair)	9kb upstream
ATL28	Ta_19900	AT2G35420	RING-H2 finger protein (protein modification)	3kb upstream

Fig 3. Genome-wide association analyses for genetic control of average DNA methylation. We show only the results for intergenic methylation; for full results see Fig S3. (A) Manhattan plots, with the top variants labelled with the neighbouring genes potentially affecting methylation. The genome-wide significance (horizontal red lines), was calculated based on unlinked variants as in Sobota et al. (2015) (46), the suggestive-line (blue) corresponds to –log(p)=5. (B) Corresponding to each Manhattan plot on the left, enrichment of *a priori* candidates and expected false discovery rates (both as in Atwell et al. 2010 (47)) for stepwise significance thresholds. (C) The allelic effects of the red-marked variants in the corresponding Manhattan plots on the left, with genotypes on the x-axes and the average methylation on the y-axes. (D) The candidate genes marked in panel A, their putative functions and distances to the top variant of the neighbouring peaks. Bold font indicates *a priori* candidates that were included in the enrichment analyses.

Methylation relationships with climate of origin

To test for environmental associations of methylation variation, we compiled bioclimatic data (see Methods section for details) for our 36 study populations and analysed the relationships between climatic variables and the average methylation in different sequence contexts and genomic features, correcting for population structure with the same IBS matrix used in the GWA analyses. We found that average methylation was positively correlated with several climate variables reflecting variation in mean temperatures, but negatively with variables related to temperature variability, such as the mean diurnal range and annual temperature range (Fig 4). In other words, plants originating from colder origins or such with more fluctuating temperature environments had lower overall methylation. In contrast to the temperature variables, methylation was not associated with the precipitation variation of the population of origin, and there was also little association with latitude (Fig 4). The latter at first appears counterintuitive, because latitude is usually correlated with temperature, but in our case latitude is confounded with altitude – more southern samples were collected at higher elevations (Table S1) – and therefore poorly correlated with temperature.

The described climate-methylation associations were generally strongest in CHG and CHH contexts, particularly for methylation that occurred in promoters or gene bodies. Accordingly, these combinations of sequence contexts and genomic features clustered together based on similarity of climate-associations, whereas almost all mCG genomic features formed a separate cluster, and so did intergenic and TE-related methylation variation in CHG and CHH, respectively (Fig 4). The only exception from these patterns was mCG in gene bodies, which had climate associations similar to mCHH, and the mCG in TEs and intergenic regions, which was the only type of methylation not negatively associated with temperature variability (Fig 4).

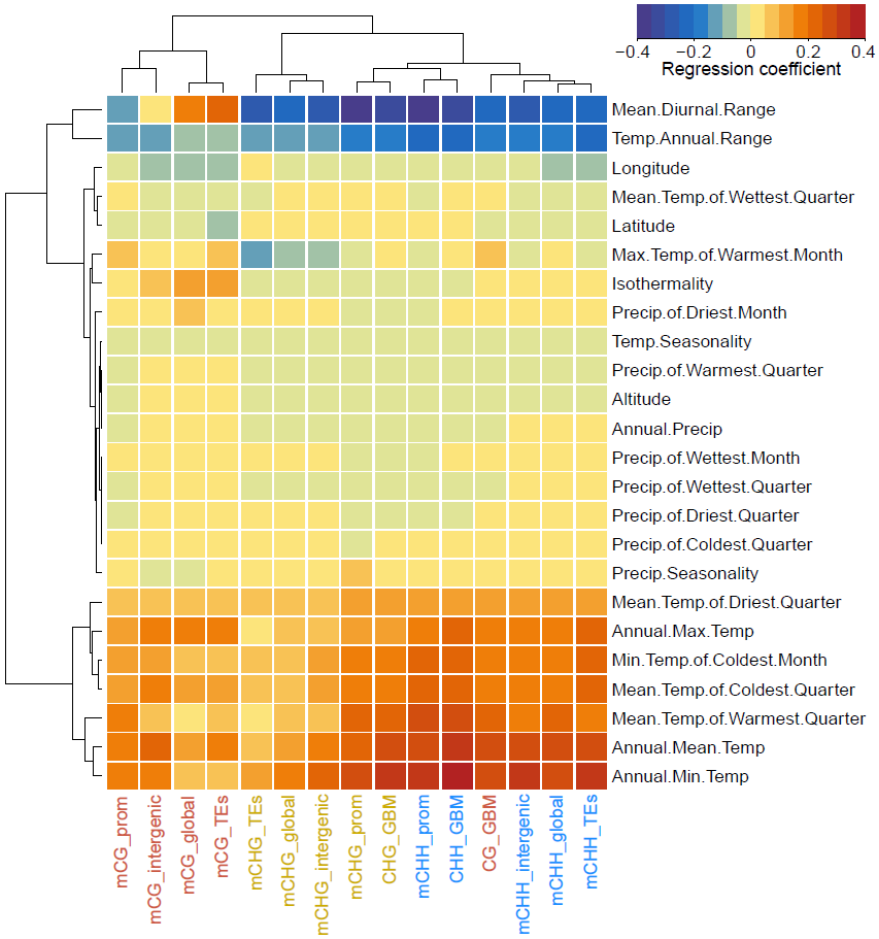


Fig 4. Climate-methylation associations. A Heatmap of the correlations between average methylation and different climatic variables (rows: Precip: precipitation; Temp: temperature), separately for different sequence contexts and genomic features (columns: prom: promoter; GBM: gene body methylation). Both rows and columns are clustered by their multivariate similarity in association patterns.

DMR variance decomposition

Having established associations of methylation variation with genetic background and environment of origin, we sought to investigate the relative importance of these two drivers in our study system, and how this might vary between sequence contexts and genomic features. To address these questions, we analysed methylation variation at the level of DMRs. We identified around 44k DMRs in CG, 12k DMRs in CHG and 77k DMRs in CHH (see Methods for details on the DMR calling), and quantified their overlap with different genomic features. Most DMRs were located in TEs, and decreasing numbers in intergenic regions, promoters and gene bodies (Figure 5B).

To quantify the degrees of genetic versus environmental determination, we then analysed three mixed models for each DMR that included either a distance matrix based on genetic variants in *cis*, on genetic variants in *trans*, or on multivariate climatic distances. Across all DMRs, genetic similarity based on *trans*-variants explained the largest proportions of methylation variance in all contexts (Fig

5A). Most variance was explained in CHG-DMRs, followed closely by CG-DMRs, but in CHH-DMRs the amounts of variance explained were generally much lower. Interestingly, the explanatory power of environmental variation relative to that of genetic variation gradually increased from the more stable mCG towards the less stable mCHG and mCHH (Fig 5A and C).

Although genetic variation in *trans* was on average the strongest predictor of methylation variance, there were large differences between individual DMRs, and we observed that sometimes genetic variation in *cis* or climatic distance, too, could be the strongest predictor. To study this more systematically, we classified all DMRs based on their strongest predictor, and we found that the fraction of DMRs in which climate was a stronger predictor of methylation variance than any of the genetic distances increased from CG to CHG to CHH (Fig 5C). In CHH, 25-30% of all DMRs had climatic distance as their strongest predictor. To find out if *cis*-, *trans*- and climate-predicted DMRs were enriched close to genes responsible for different functions, we ran separate GO enrichment analyses for the genes neighbouring these three classes of DMRs. However, only for the *trans*-predicted DMRs we found significant enrichment of a few GO terms (Fig S4), while there were none for the other two DMR classes.

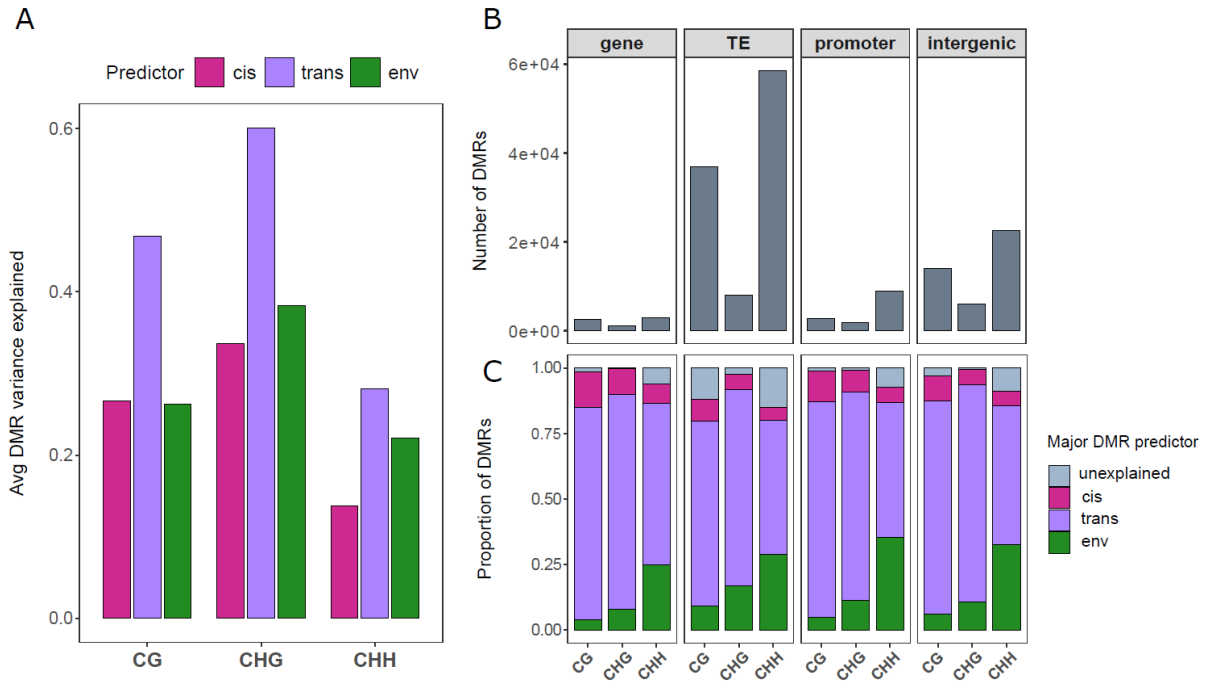


Fig 5. Genetic versus environmental predictors of DMR variance. (A) The variance in DMR methylation explained by genetic similarity in *cis*, genetic similarity in *trans* and climatic similarity, averaged across all DMRs. (B) The numbers of DMR identified in different genomic features and sequence contexts, and (C) the fractions of these individual DMRs where *cis*-variation, *trans*-variation or climatic variation are the major predictors. DMRs where none of the three predictors explained >10% of the variance are classified as "unexplained".

Discussion

Understanding natural epigenetic variation requires combining large-scale surveys of natural populations with high-resolution genomics and environmental data. Here, we studied European populations of *T. arvense* to assess how climate of origin and genetic background shaped their heritable DNA methylation variation. We found epigenetic population structure and confirmed the genomic patterns of methylation of the *T. arvense* genome (36) in a large natural collection. Most importantly, both genetic background and climate of origin were significantly associated with methylation variation, but their relative predictive power varied depending on DNA sequence context.

Our analysis of population structure detected two main clades, one composed of lines from all surveyed countries and a smaller one with almost exclusively lines from Sweden. A latitudinal gradient was also clear within the larger clade. The epigenetic population structure generally resembled the genetic one, with decreasing degrees of similarity from CG to CHG and to CHH sequence contexts (Fig 1B). These differences between contexts might reflect their different stability, caused by differences in the maintenance machineries (13,14) and possibly different proportions of genetic versus environmental control. Moreover, mCG shows stronger geographic patterns in genes and promoters than in TEs, possibly indicating a higher stability or selection for this kind of methylation (Fig S1).

Across all lines, we calculated a global average methylation of 18.5%, which is fairly high in the Brassicaceae family (20), particularly in comparison to *A. thaliana* (5.8%) (12). The high global methylation is related to the high TE content of the *T. arvense* genome (~60%) Nunn et al. 2021), but also to a higher CHH methylation (13.2% across all lines) than it is known for most other angiosperms (20). The levels of CG and CHG methylation (53.7% and 15.7% across all lines), in contrast, are more similar to other Brassicaceae (20). As expected, we found that methylation was very unevenly distributed not only between sequence contexts, but also between genomic features, with high levels of methylation particularly in CG context, and in TEs and intergenic regions (Fig 2A). Gene body methylation was generally very low, with some 90% of the genes nearly unmethylated in all contexts, and the results were similar, albeit much less extreme, for promoters (Fig 2B)(36). This uncommonly low GBM is present in other Brassicaceae (20), in particular in the close relative *Eutrema salsugineum* and might have evolved before speciation between Thlaspi and Eutrema (36). Moreover, genes methylated only in CG were enriched for defence mechanisms and not for housekeeping functions (Fig S2) as in many other plant species (20). On the contrary, TE-like methylated genes (methylated in all contexts), which are usually pseudogenes, were enriched for some constitutive functions so

probably include a portion of housekeeping genes (Fig S2). Overall these findings suggest that GBM in *T. arvense* differs relevantly from previously studied plant species (20).

To understand the genetic basis of methylation variation in *T. arvense*, we used GWA analyses, testing for associations between DNA sequence variation and average methylation levels in different sequence contexts and genomic features. With a strict Bonferroni correction, we did not detect any significant genetic variants, which probably resulted from a combination of our moderate number of only 207 sequenced lines, the nested sampling design, and the high number of tests (compared to *A. thaliana*) in a ~500 Gb genome. However, for some methylation phenotypes, we found strong enrichment of *a priori* candidates neighbouring genes known to play a role in DNA methylation from *A. thaliana* studies, and this indicates that many of our top peaks are likely to be true positives (Fig 3 and S3). Examples include the peaks detected next to the genes *AGO9*, *DRM3* and *LIG1*, which are all part of the DNA methylation machinery of *A. thaliana* (13,14), and which were also among our *a priori* candidates (Table S3). In addition to these 'expected' candidates, we found several additional peaks next to genes that were indirectly linked to DNA methylation, with predicted functions such as histone acetylation, DNA repair and ubiquitination (Table S4). The latter in particular is a post-translational modification which was previously shown to affect methylation in several ways (28–33). These new candidate genes were not in our *a priori* list, which explains the drop of enrichment at high -log(p) in several GWA analyses (Fig 3B and S3). Our results show that while there appears to be partial overlap in the genetic control of DNA methylation between *T. arvense* and *A. thaliana*, there are also important differences. Some of our strongest candidates have not been associated with DNA methylation before, particularly not in natural populations. Functional characterization of these "new" candidates will be necessary to confirm our findings and understand the mechanisms of action of these genes.

Finally, some interesting associations warrant further exploration and could uncover functional differences with *A. thaliana* in the methylation machineries of different sequence contexts. For example we find a peak for mCG, next to a *DRM3* orthologue, involved in non-CG methylation maintenance in Arabidopsis (43–45), and vice versa a peak for mCHH of promoters and TEs right next (3kb upstream) to an orthologue of the mCG maintenance methyltransferase *DMT1*. On the contrary, the high similarity between mCHG and mCHH in regard to their genetic basis, as shown by the strong overlap of GWA results, seems to be a common feature in the plant kingdom (13,14).

Natural epigenetic variation was not only associated with genetic background in our study, but also with climate of origin. These correlations were generally much stronger than those with latitude or longitude, which supports the idea that the observed correlations reflect adaptive processes and not just the combination of epigenetic drift and isolation-by-distance. Specifically, we found average

methylation to be positively correlated with mean temperature but negatively with temperature variability (Fig 4). Our field survey particularly captured the cold end of the distribution range of *T. arvense* (Mean Annual Temp. 6.5 – 11.1 °C). Previous studies showed that cold can induce DNA demethylation in plants (48–50) and that demethylation in turn can be associated with expression of cold-resistance genes and increased freezing tolerance (51,52). The observed negative correlations between methylation and temperature might therefore reflect adaptation to cold and to capturing the cold end of the distribution. This interpretation is further supported by the fact that correlations with minimum temperatures were generally stronger than with bioclimatic variables capturing maximum temperature (Fig 4) and explains why a similar study found negative correlations between temperature and methylation in Arabidopsis accessions sampled on a range including many warmer locations (12). The negative relationship between DNA methylation and temperature variability (Mean Diurnal Range and Temperature Annual Range) is more challenging to interpret, as there have so far been no experimental tests manipulating environmental variability in temperature. However, lower DNA methylation is often associated with lower genome stability (53,54), and it is conceivable that in fluctuating and thus less predictable environmental conditions, lower genome stability and higher transposon activity could be adaptive. Finally, we did not find any association between DNA methylation and the precipitation of the population origins. However, this may largely be a result of our latitudinal sampling design, which maximized temperature but not precipitation variation. None of our sampling sites were particularly dry or particularly wet/oceanic (Annual Prec. 475 – 869 mm).

To better understand the predictive power of climate of origin versus genetic background, we finally analysed the variance in methylation levels of individual DMRs. We found that, across all DMRs, genetic variation in *trans* generally explained more DMR variation than climatic variation or genetic variation in *cis*. However, there was a trend from CG to CHG to CHH that the explanatory power of climate increased relative to that of genetic background (Fig 5A). In CHH, climate was the strongest predictor of methylation variation in over one quarter of the individual DMRs; in promoters this was true for even 35% of the DMRs (Fig 5B). These results further support the idea that methylation variation, particularly in CHG and CHH, is not only involved in plant responses to short-term stress (17) but also in longer-term environmental adaptation. Moreover, the observation that sometimes climate was the strongest predictor, indicates that at least part of the climate-methylation associations could be independent of DNA sequence variation (5). Clearly, further work is needed to support these speculations, in particular high-resolution analyses that disentangle the genomic versus epigenomic basis of relevant phenotypes related to climatic tolerances. We attempted to get some hints of the functional basis of the observed genomic-methylation and climate-methylation relationships by analysing GO enrichment in the neighbouring genes of *trans*-, *cis*- and

environmentally-associated DMRs, and we found some enrichment, mostly related to housekeeping functions, for *trans*-DMRs, but none for *cis*- and environmentally-associated DMRs (Fig S4). However, the functional annotation had GO terms for only less than half of our candidate genes, so our GO enrichment analysis had rather limited power.

In summary, our study is the first large-scale investigation of DNA methylation variation in natural plant populations beyond the Arabidopsis model. We found that *T. arvense* natural DNA methylation variation is shaped by genetic and environmental factors, and that the relative contributions of the two drivers vary strongly between sequence contexts. Methylation variation in CG is generally the most similar to, and best predicted by, genetic variation. Moving to CHG and particularly CHH, the genetic determination decreases making environmental determination relatively higher. Our results thus indicate that DNA methylation could play a role in the large-scale environmental adaptation of *T. arvense*. Further experimental research, in particular dissecting adaptive phenotypes, is necessary to corroborate this hypothesis. There are currently efforts underway to develop *T. arvense* into a new biofuel and winter cover crop (37–41), and any insights into the genomic basis of climate and other environmental adaptation will be highly relevant to these efforts, particularly to deal with future climates.

Materials and Methods

Sampling and plant growth

In July 2018, we collected *T. arvense* seeds from 36 natural populations in six European regions, spanning from southern France to central Sweden, and used them to conduct a common environment experiment in Tübingen, Germany. The experiment started at the end of August 2019 and lasted about two months. Upon sowing 207 lines in 9x9 cm pots filled with soil, we stratified them for 10 d at 4°C in the dark. We then transferred the seeds to a glasshouse and transplanted seedlings to individual pots upon germination. The glasshouse had a 15/9 h light/dark cycle (6 a.m. to 9 p.m.) with temperature and humidity conditions averaging 18°C and 30% at night and 22°C and 25% during the day. External conditions influenced these parameters, resembling natural growing conditions. 46 d after the end of the stratification period, we collected the 3rd or 4th true leaf and snap-froze it in liquid nitrogen.

Library preparation and sequencing

Using the DNeasy Plant Mini Kit (Qiagen, Hilden, DE), we extracted DNA from disrupted leaf tissue obtained from the 3rd or 4th true leaf. For each sample, we sonicated (Covaris) 300 ng of genomic DNA to a mean fragment size of ~350 bp and used the resulting DNA for both genomic and bisulfite

libraries. The NEBNext® Ultra™ II DNA Library Prep Kit for Illumina was used for library preparation and was combined with EZ-96 DNA Methylation-Gold™ MagPrep (ZYMO) for bisulfite libraries. Briefly, the procedure involved: i) end repair and 3' adenylation of sonicated DNA fragments, ii) NEBNext adaptor ligation and U excision, iii) size selection with AMPure XP Beads (Beckman Coulter, Brea, CA), iv) splitting DNA for bisulfite (2/3) and genomic (1/3) libraries, v) bisulfite treatment and cleanup of bisulfite libraries, vi) PCR enrichment and index ligation using Kapa HiFi Hot Start Uracil+ Ready Mix (Agilent) for bisulfite libraries (14 cycles) and NEBNext Ultra II Q5 Master Mix for genomic libraries (4 cycles), vii) final size selection and cleanup. Finally, we sequenced paired-end for 150 cycles. Genomic libraries were sequenced on Illumina NovaSeq 6000 (Illumina, San Diego, CA), while bisulfite libraries were sequenced on HiSeq X Ten (Illumina, San Diego, CA).

Variant calling, filtering and imputation

Base calling and demultiplexing of raw sequencing data were performed by Novogene using the standard Illumina pipeline. After quality and adaptor trimming using cutadapt v2.6 (M. Martin 2011), we aligned reads to the reference genome with BWA-MEM v0.7.17 (55). We then performed variant calling with GATK4 v4.1.8.1 (56,57) following the best practices for Germline short variant discovery (https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels-). Briefly, we marked duplicates with MarkDuplicatesSpark and ran HaplotypeCaller to obtain individual sample GVCF files. We combined individual GVCF files running GenomicsDBImport and GenotypeGVCFs successively and parallelizing by scaffold, obtaining single-scaffold multisample vcf files. We then re-joined these files with GatherVcfs. Upon assessment of quality parameters distributions, we removed low quality variants using VariantFiltration with different filtering parameters for SNPs (QD < 2.0 || SOR > 4.0 || FS > 60.0 || MQ < 20.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0) and other variants (QD < 2.0 || QUAL < 30.0 || FS > 200.0 || ReadPosRankSum < -20.0). Using vcftools v0.1.16 (58), we further filtered scaffolds with less than three variants and variants with multiple alleles or more than 10% missing values. Prior to imputation, we only applied a mild Minor Allele Frequency (MAF) > 0.01 filtering not to reduce imputation accuracy (59). Imputation with BEAGLE 5.1 (60) recovered the few missing genotype calls left, outputting a complete multisample vcf file.

Methylation analysis

The EpiDiverse WGBS pipeline is specifically designed for bisulfite reads mapping and methylation calling in non-model species (https://github.com/EpiDiverse/wgbs) (61). We used it to perform quality control (FastQC), base quality and adaptor trimming (cutadapt), bisulfite aware mapping (erne-bs5), duplicates detection (Picard MarkDuplicates), alignment statistics and methylation calling

(Methyldackel). In the mapping step, we only retained uniquely-mapping reads longer than 30bp. The pipeline outputs context-specific (CG, CHG and CHH) individual-sample bedGraph files, which we filtered for coverage and combined in multisample unionbed files using custom scripts and bedtools (62). We retained all cytosines with coverage > 3 in at least 75% of the samples and calculated global and CG, CHG and CHH average methylation.

To extract the average methylation of genomic features, we intersected (bedtools) (62) unionbed files with genomic features (genes, TEs, promoters and intergenic regions) and averaged methylation of all intersected cytosines. We also extracted methylation of individual genes, promoters and TEs across all samples and plotted their distributions. We then used this information to calculate the average methylation of genes, excluding unmethylated ones (average mC < 5% across all lines) and used it for GWA. For PCA, we used the R (63) function prcomp(). Genome wide PCAs were only based on positions without missing values as these were already a large amount (always > 1 million). Instead when restricting to genomic features we allowed for 2% NAs and imputed these with the missMDA R package (64) to include a larger amount of positions (always > 0.8 million). Nevertheless comparison of PCA plots with and without imputation gave very similar results.

Population genetic and GWA analysis

For basic genetic population structure analysis, including PCA plots and generation of the IBS matrix, we applied a mild MAF filtering (MAF>0.01) and performed variants pruning with PLINK v1.90b6.12 (65), using a window of 50 variants, sliding by five and a maximum LD of 0.8. Upon this filtering, we also used PLINK to generate the IBS matrix used in several analyses to correct for population structure or for DMRs variance decomposition. For PCA, we used the R (63) function prcomp().

We ran GWA analysis for multiple phenotypes using a custom script based on the R package rrBLUP (66), which allows to run mixed models correcting for population structure with the above-mentioned IBS matrix. We used biallelic variants and applied a MAF > 0.4 cutoff. For Manhattan and QQplots we used the "qqman" package (67), calculating the genome-wide significance threshold according to Sobota et al. (2015) (46). We run GWA analysis using each average methylation context (CG, CHG and CHH) feature (global, gene bodies, TEs, promoters and intergenic regions) combination as phenotype. For genes we also calculated average methylation of methylated genes, excluding unmethylated ones (average GBM across all lines > 5%), ending up with a total of 18 phenotypes (Table S2). We occasionally observed a positive correlation between average methylation and coverage across lines, probably due to library preparation bias. In these cases we fit a linear model to the data using the logarithm of coverage (from bam files), as this gave the best fit in all cases, and used the residuals for GWA analysis. Finally, we transformed the average methylation phenotypes

that deviated strongly from normality with the Gaussianize function from the R package "LambertW" (68) or by Inverse Normal Transformation when the former was not obtaining normality. A list of all phenotypes used and corrections and transformations applied, can be found in Table S2.

With the double aim of validating GWA results and comparing with previous *A. thaliana* studies, we performed enrichment of variants neighbouring *a priori* candidate genes, according to the method established by Atwell et al. (2010) (47). We made a few additions to the methylation candidate gene list used by Kawakatsu et al. (2016) (12), kindly provided by the authors, extracted all *T. arvense* orthologues that we could retrieve from orthofinder (69) analysis and used them for our *a priori* candidate genes list (Table S3). Briefly, we attributed "*a priori* candidate" status to all variants within 20kb from genes in the list and calculated enrichment for increasing -log(p) thresholds as the ratio between Observed frequency (sign. *a priori* candidates/sign. variants) and Background frequency (total *a priori* candidates/total variants). Using the same formula adopted by Atwell et al. (2010) (47), we additionally calculated an upper bound for the FDR among candidates.

Climate-methylation correlations

To obtain bioclimatic variables for the 25 years predating the experiment, we started by downloading temperature and precipitation daily gridded data from the Copernicus database (v21.0) (70). All downstream analysis were conducted in R (63). We subsetted our population locations with the "ncdf4" package (71), calculated monthly averages and extracted bioclimatic variables with "dismo" (72). Finally, we averaged bioclimatic variables from 1994 to 2018, the year of collection. To test for climatic patterns in methylation, we ran mixed models for all average methylation vs. bioclimatic variables combinations, using the relmatLmer() function from the R package "lme4qtl" (73) and correcting for population structure using the same IBS matrix used for GWA analysis.

DMR calling

The EpiDiverse toolkit (61) includes a DMR pipeline based on metilene (74), which calls DMRs between all possible pairwise comparisons between user-defined groups. We used this tool to call DMRs using the 36 populations as groups, a minimum coverage of five (cov > 4) and default values for all other parameters. We complemented the pipeline with a custom downstream workflow to obtain DMRs for the whole collection from comparison-specific DMRs. Briefly, since the pipeline output had an enrichment of short and close DMRs (particularly in CHH), we joined all comparison-specific DMRs that were closer than 146bp and had the same directionality (higher methylation in the same group). 146bp was chosen for consistency with the pipeline fragmentation parameter. We then merged DMRs from all pairwise comparisons (bedtools) (62) in a unique file and re-extracted

average methylation of the resulting regions from all samples. Finally, we filtered DMRs with a minimum methylation difference of 20% (CG) or 15% (CHG and CHH) in at least 5% of the samples. This ensured to select DMRs with variability at the level of the whole collection.

DMR variance decomposition

To quantify the variance in methylation explained by *cis*-variants, *trans*-variants, and by the environment, we ran three mixed models for each individual DMR using the marker_h2() function from the R package "heritability" (75). Each model had one random factor matrix, capturing one of the three predictors. For *cis* we used an IBS matrix generated with PLINK v1.90b6.12 (65) from variants within 50kb from the DMR middle point. For *trans* we used the same IBS matrix used for all other analysis, described in the previous chapter. For the environment we calculated the Euclidean distance between locations, based on all Bioclimatic Variables averaged over 25 years before the sampling (1994-2018), and further reversed and normalized the matrix to obtain a similarity matrix in a 0 to 1 range. To summarize the results we: i) averaged *cis*, *trans* and environment explained variance across all DMRs and II) classified each DMR based on the mayor predictor.

Supporting Information

S1 Fig. PCA plots of all 207 lines. (A) Complement to Fig 1B with latitude-coloured PCA plots for the missing PC. (B) latitude-coloured PCA plots based on methylation of specific genomic features (genes, TEs and promoters).

S2 Fig. Genes methylated in each context and GO enrichment analysis. Cutoffs of mCG > 20% and mCHG/mCHH > 8%, across all lines, were used to identify methylated genes (values chosen based on the distributions in Fig2B). (A) Venn diagram of number of genes methylated in each context which were also used for the GO enrichment. Genes methylated only in CG are ladled as "gbM", genes methylated in all contexts as "TE-like". (B) GO enrichment analysis of methylated genes. Only significant results for GO terms with minimum gene count of four are reported.

S3 Fig. Complete methylation GWA results. Manhattan plots, enrichment of *a priori* candidate variants and QQplots for all average methylation phenotypes. more5met: average methylation of genes with GBM > 5% across all lines. . The genome-wide significance (horizontal red lines), was calculated based on unlinked variants as in Sobota et al. (2015) (46), the suggestive-line (blue) corresponds to –log(p)=5. Top variants are labelled with the neighbouring genes potentially affecting methylation.

S4 Fig. GO enrichment analysis of genes neighbouring trans-DMRs. Genes neighbouring (2kb max) *cis*, *trans* and *env*-DMRs were used for individual GO term enrichment analysis, but only the *trans*-DMRs gene set was enriched for any significant term.

S1 Table. Geographic locations of all *T. arvense* populations. Geographic coordinates, elevation and size of all populations.

S2 Table. List of all average methylation variables, coverage corrections and transformations applied. Coverage correction indicates that, prior to GWA, residuals were extracted from a linear model with log(coverage) as predictor. INT indicates Inverse Normal Transformation. Gaussianize function was applied using the "LambertW" R package (68). more5met: GBM of genes with average methylation > 5% across all lines.

S3 Table. List of *Thlaspi arvense a priori* candidate genes. *T. arvense* genes and the respective *A. thaliana* orthologues with known roles in methylation. We used this list for the enrichment of a priori candidate variants performed upon GWA.

S4 Table. GWA candidate genes. List of all genes located within 15kb from variants significant to -log(p) > 5, including average methylation phenotypes where the association was found, *a priori* candidate status and relevant functional putative roles. Genes with predicted function possibly affecting methylation are highlighted in bold.

Germplasm, data and code availability

Seed material from the sequenced lines is available at the Nottingham Arabidopsis Stock Centre (NASC) under stock numbers N950001 to 950204. Genomic and bisulfite sequencing raw data are available on the ENA Sequence Read Archive (www.ebi.ac.uk/ena/) under study accession number PRJEB50950. Reference genome and annotations were previously published by Nunn et al. (2021)(36). From the annotation, we subset confident de novo gene annotations with an annotation edit distance score < 1.0 (source:T_arvense_v2). Average methylation values extracted from sequencing data and GWA results (-log(p)>1) in a format compatible with the Integrative Genomics Viewer (https://software.broadinstitute.org/software/igv/) are available on Zenodo (10.5281/zenodo.6361978).

All the code used in this study is available and documented on Github. Pipelines for methylation and DMR calling from WGBS data are on the EpiDiverse Github (https://github.com/EpiDiverse). The workflow for downsteam analysis of methylation data is on https://github.com/Dario-Galanti/WGBS_downstream. Scripts for downstream processing of DMRs and DMRs variance decomposition are on https://github.com/Dario-Galanti/popDMRs_refine_VCA. Scripts for variant calling, filtering and imputation, performed on the Baden-Wurttenberg BinAC cluster, are on

https://github.com/Dario-Galanti/BinAC_varcalling. Finally, scripts for running GWA analysis and the enrichment of *a priori* candidates are available at https://github.com/Dario-Galanti/multipheno_GWAS.

Funding

This work is part of the European Training Network EpiDiverse (https://epidiverse.eu), which received funding from the EU Horizon 2020 program under Marie Skłodowska-Curie grant agreement No 764965. This work was also supported by the European Union's Horizon 2020 research and innovation program via the European Research Council (ERC) Grant agreement No. 716823 "FEAR-SAP" to C.B., by the Austrian Academy of Sciences and by the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG.

Acknowledgments

We thank the whole EpiDiverse network for the amazing support and productive discussions, in particular Adrián Contreras-Garrido, for providing orthofinder results and discussing analysis, and Bárbara Díez Rodríguez and Iris Sammarco for really useful discussions and suggestions. We also thank Detlef Weigel for suggestions on data analysis and Magnus Nordborg and Eriko Sasaki for feedback on GWA analysis and for sharing their list of *a priori* candidate genes. Finally, we thank Anupoma Troyee and Valentina Vaglia for the help with sampling, Sabine Silberhorn, Christiane Karasch-Wittmann, Eva Schloter, Julia Rafalski and Elodie Kugler for help with the greenhouse experiment, and Katharina Jandrasits for help with the sequencing library preparation. For computing, we acknowledge Prof. Peter Stadler at the University of Leipzig and David Langenberger from ecSeq, for hosting the EpiDiverse servers. We also acknowledge the High Performance and Cloud Computing Group at the Zentrum für Datenverarbeitung of the University of Tübingen for managing the BinAC server.

Author Contributions

Conceptualization: Oliver Bossdorf, J.F. Scheepens and Claude Becker.

Funding acquisition: Oliver Bossdorf and Claude Becker.

Sampling: Dario Galanti and Daniela Ramos.

Wet lab: Dario Galanti and Daniela Ramos.

Sequencing data generation and analysis support: Adam Nunn, Isaac Jorge Rodriguez and Claude Becker.

Data analysis and visualization: Dario Galanti.

Supervision: Oliver Bossdorf and J.F. Scheepens.

Writing: Dario Galanti, Oliver Bossdorf.

Writing – review & editing: All the authors.

References

1. Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, et al. Transgenerational epigenetic instability is a source of novel methylation variants. Science. 2011 Oct 21;334(6054):369–73.

2. Becker C, Hagmann J, Müller J, Koenig D, Stegle O, Borgwardt K, et al. Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. Nature. 2011 Dec;480(7376):245–9.

3. Lämke J, Bäurle I. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. Genome Biology. 2017 Jun 27;18(1):124.

4. He Y, Li Z. Epigenetic Environmental Memories in Plants: Establishment, Maintenance, and Reprogramming. Trends in Genetics [Internet]. 2018 Aug 22 [cited 2018 Aug 30]; Available from: http://www.sciencedirect.com/science/article/pii/S0168952518301276

5. Richards CL, Alonso C, Becker C, Bossdorf O, Bucher E, Colomé-Tatché M, et al. Ecological plant epigenetics: Evidence from model and non-model species, and the way forward. Ecol Lett. 2017 Dec 1;20(12):1576–90.

6. Schmid MW, Heichinger C, Schmid DC, Guthörl D, Gagliardini V, Bruggmann R, et al. Contribution of epigenetic variation to adaptation in Arabidopsis. Nature Communications. 2018 Oct 25;9(1):4446.

7. Münzbergová Z, Latzel V, Šurinová M, Hadincová V. DNA methylation as a possible mechanism affecting ability of natural populations to adapt to changing climate. Oikos. 2019;128(1):124–34.

8. Paun O, Bateman RM, Fay MF, Hedrén M, Civeyrel L, Chase MW. Stable Epigenetic Effects Impact Adaptation in Allopolyploid Orchids (Dactylorhiza: Orchidaceae). Mol Biol Evol. 2010 Nov 1;27(11):2465–73.

9. Lira-Medeiros CF, Parisod C, Fernandes RA, Mata CS, Cardoso MA, Ferreira PCG. Epigenetic Variation in Mangrove Plants Occurring in Contrasting Natural Environment. PLOS ONE. 2010 Apr 26;5(4):e10326.

10. Gugger PF, Fitz-Gibbon S, Pellegrini M, Sork VL. Species-wide patterns of DNA methylation variation in Quercus lobata and their association with climate gradients. Molecular Ecology. 2016 Apr 1;25(8):1665–80.

11. 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 [cited 2018 May 8];4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4413256/

12. Kawakatsu T, Huang SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, et al. Epigenomic Diversity in a Global Collection of Arabidopsis thaliana Accessions. Cell. 2016 Jul 14;166(2):492–505.

13. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics. 2010 Mar;11(3):204–20.

14. Zhang H, Lang Z, Zhu J-K. Dynamics and function of DNA methylation in plants. Nature Reviews Molecular Cell Biology. 2018 May 21;1.

15. Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, Matzke M, et al. Role of the DRM and CMT3 Methyltransferases in RNA-Directed DNA Methylation. Current Biology. 2003 Dec 16;13(24):2212–7.

16. 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 Jan;21(1):64–72.

17. Liu J, He Z. Small DNA Methylation, Big Player in Plant Abiotic Stress Responses and Memory. Frontiers in Plant Science. 2020 [cited 2022 Mar 9];11. Available from: https://www.frontiersin.org/article/10.3389/fpls.2020.595603

18. Schmitz RJ, Schultz MD, Urich MA, Nery JR, Pelizzola M, Libiger O, et al. Patterns of population epigenomic diversity. Nature. 2013 Mar;495(7440):193–8.

19. Seymour DK, Koenig D, Hagmann J, Becker C, Weigel D. Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization. PLoS Genet. 2014 Nov;10(11):e1004785.

20. Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, et al. Widespread natural variation of DNA methylation within angiosperms. Genome Biology. 2016 Sep 27;17:194.

21. Schmitz RJ, Lewis ZA, Goll MG. DNA Methylation: Shared and Divergent Features across Eukaryotes. Trends in Genetics. 2019 Nov 1;35(11):818–27.

22. Hazarika RR, Serra M, Zhang Z, Zhang Y, Schmitz RJ, Johannes F. Molecular properties of epimutation hotspots. Nat Plants. 2022 Jan 27;1–11.

23. Cubas P, Vincent C, Coen E. An epigenetic mutation responsible for natural variation in floral symmetry. Nature. 1999 Sep;401(6749):157–61.

24. 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. Nature Genetics. 2006 Aug;38(8):948–52.

25. Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, et al. A transposon-induced epigenetic change leads to sex determination in melon. Nature. 2009 Oct;461(7267):1135–8.

26. Cortijo S, Wardenaar R, Colomé-Tatché M, Gilly A, Etcheverry M, Labadie K, et al. Mapping the Epigenetic Basis of Complex Traits. Science. 2014 Mar 7;343(6175):1145–8.

27. Sasaki E, Kawakatsu T, Ecker JR, Nordborg M. Common alleles of CMT2 and NRPE1 are major determinants of CHH methylation variation in Arabidopsis thaliana. PLOS Genetics. 2019 dic;15(12):e1008492.

28. Bostick M, Kim JK, Estève P-O, Clark A, Pradhan S, Jacobsen SE. UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. Science. 2007 Sep 21;317(5845):1760–4.

29. Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature. 2007 Dec;450(7171):908–12.

30. Kraft E, Bostick M, Jacobsen SE, Callis J. ORTH/VIM proteins that regulate DNA methylation are functional ubiquitin E3 ligases. The Plant Journal. 2008;56(5):704–15.

31. Kim J, Kim JH, Richards EJ, Chung KM, Woo HR. Arabidopsis VIM Proteins Regulate Epigenetic Silencing by Modulating DNA Methylation and Histone Modification in Cooperation with MET1. Molecular Plant. 2014 Sep 1;7(9):1470–85.

32. Chen J, Liu J, Jiang J, Qian S, Song J, Kabara R, et al. F-box protein CFK1 interacts with and degrades de novo DNA methyltransferase in Arabidopsis. New Phytologist. 2021;229(6):3303–17.

33. Wang J, Qiu Z, Wu Y. Ubiquitin Regulation: The Histone Modifying Enzyme's Story. Cells. 2018 Aug 27;7(9):118.

34. Gáspár B, Bossdorf O, Durka W. Structure, stability and ecological significance of natural epigenetic variation: a large-scale survey in Plantago lanceolata. New Phytologist. 2018 [cited 2018 Oct 12];0(0). Available from: https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/nph.15487

35. 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 [cited 2018 Feb 26];6. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2015.00004/full

36. Nunn A, Rodríguez-Arévalo I, Tandukar Z, Frels K, Contreras-Garrido A, Carbonell-Bejerano P, et al. Chromosome-level Thlaspi arvense genome provides new tools for translational research and for a newly domesticated cash cover crop of the cooler climates. 2021 Aug [cited 2021 Sep 1] p. 2021.07.30.454478. Available from: https://www.biorxiv.org/content/10.1101/2021.07.30.454478v1

37. Tsogtbaatar E, Cocuron J-C, Sonera MC, Alonso AP. Metabolite fingerprinting of pennycress (Thlaspi arvense L.) embryos to assess active pathways during oil synthesis. J Exp Bot. 2015 Jul;66(14):4267–77.

38. Dorn KM, Fankhauser JD, Wyse DL, Marks MD. A draft genome of field pennycress (Thlaspi arvense) provides tools for the domestication of a new winter biofuel crop. DNA Res. 2015 Apr 1;22(2):121–31.

39. Claver A, Rey R, López MV, Picorel R, Alfonso M. Identification of target genes and processes involved in erucic acid accumulation during seed development in the biodiesel feedstock Pennycress (Thlaspi arvense L.). Journal of Plant Physiology. 2017 Jan 1;208:7–16.

40. Chopra R, Johnson EB, Daniels E, McGinn M, Dorn KM, Esfahanian M, et al. Translational genomics using Arabidopsis as a model enables the characterization of pennycress genes through forward and reverse genetics. The Plant Journal. 2018;96(6):1093–105.

41. Chopra R, Johnson EB, Emenecker R, Cahoon EB, Lyons J, Kliebenstein DJ, et al. Progress toward the identification and stacking of crucial domestication traits in pennycress. Plant Biology; 2019 Apr [cited 2019 Apr 19]. Available from: http://biorxiv.org/lookup/doi/10.1101/609990

42. Geng Y, Guan Y, Qiong L, Lu S, An M, Crabbe MJC, et al. Genomic analysis of field pennycress (Thlaspi arvense) provides insights into mechanisms of adaptation to high elevation. BMC Biology. 2021 Jul 22;19(1):143.

43. Henderson IR, Deleris A, Wong W, Zhong X, Chin HG, Horwitz GA, et al. The De Novo Cytosine Methyltransferase DRM2 Requires Intact UBA Domains and a Catalytically Mutated Paralog DRM3 during RNA–Directed DNA Methylation in Arabidopsis thaliana. PLOS Genetics. 2010 ott;6(10):e1001182.

44. Zhong X, Hale CJ, Nguyen M, Ausin I, Groth M, Hetzel J, et al. DOMAINS REARRANGED METHYLTRANSFERASE3 controls DNA methylation and regulates RNA polymerase V transcript abundance in Arabidopsis. PNAS. 2015 Jan 20;112(3):911–6.

45. Costa-Nunes P, Kim JY, Hong E, Pontes O. The cytological and molecular role of DOMAINS REARRANGED METHYLTRANSFERASE3 in RNA-dependent DNA methylation of Arabidopsis thaliana. BMC Research Notes. 2014 Oct 14;7(1):721.

46. Sobota RS, Shriner D, Kodaman N, Goodloe R, Zheng W, Gao Y-T, et al. Addressing population-specific multiple testing burdens in genetic association studies. Ann Hum Genet. 2015 Mar;79(2):136–47.

47. Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, et al. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature. 2010 Jun;465(7298):627–31.

48. Steward N, Kusano T, Sano H. Expression of ZmMET1, a gene encoding a DNA methyltransferase from maize, is associated not only with DNA replication in actively proliferating cells, but also with altered DNA methylation status in cold-stressed quiescent cells. Nucleic Acids Research. 2000 Sep 1;28(17):3250–9.

49. Conde D, Le Gac A-L, Perales M, Dervinis C, Kirst M, Maury S, et al. Chilling-responsive DEMETER-LIKE DNA demethylase mediates in poplar bud break. Plant Cell Environ. 2017 Oct;40(10):2236–49.

50. Lai Y-S, Zhang X, Zhang W, Shen D, Wang H, Xia Y, et al. The association of changes in DNA methylation with temperature-dependent sex determination in cucumber. Journal of Experimental Botany. 2017 May 17;68(11):2899–912.

51. Xie HJ, Li H, Liu D, Dai WM, He JY, Lin S, et al. ICE1 demethylation drives the range expansion of a plant invader through cold tolerance divergence. Molecular Ecology. 2015 Feb 1;24(4):835–50.

52. Xie H, Sun Y, Cheng B, Xue S, Cheng D, Liu L, et al. Variation in ICE1 Methylation Primarily Determines Phenotypic Variation in Freezing Tolerance in Arabidopsis thaliana. Plant and Cell Physiology. 2019 Jan 1;60(1):152–65.

53. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet. 2007 Apr;8(4):272–85.

54. Zhou D, Robertson KD. Role of DNA Methylation in Genome Stability. Elsevier Inc.; 2016. p. 409–24. Available from: http://www.scopus.com/inward/record.url?scp=85022019460&partnerID=8YFLogxK

55. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009 Jul 15;25(14):1754–60.

56. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Angel G del, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Current Protocols in Bioinformatics. 2013;43(1):11.10.1-11.10.33.

57. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GAV der, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2018 Jul 24;201178.

58. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011 Aug 1;27(15):2156–8.

59. Roshyara NR, Kirsten H, Horn K, Ahnert P, Scholz M. Impact of pre-imputation SNP-filtering on genotype imputation results. BMC Genet. 2014 Aug 12;15:88.

60. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. The American Journal of Human Genetics. 2018 Sep 6;103(3):338–48.

61. Nunn A, Can SN, Otto C, Fasold M, Díez Rodríguez B, Fernández-Pozo N, et al. EpiDiverse Toolkit: a pipeline suite for the analysis of bisulfite sequencing data in ecological plant epigenetics. NAR Genomics and Bioinformatics. 2021 Dec 1;3(4):lqab106.

62. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010 Mar 15;26(6):841–2.

63. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

64. Husson F, Josse J. missMDA: Handling Missing Values with Multivariate Data Analysis. 2020 [cited 2022 Feb 1]. Available from: https://CRAN.R-project.org/package=missMDA

65. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics. 2007 Sep 1;81(3):559–75.

66. Endelman J. rrBLUP: Ridge Regression and Other Kernels for Genomic Selection. 2019 [cited 2022 Feb 1]. Available from: https://CRAN.R-project.org/package=rrBLUP

67. Turner S. qqman: Q-Q and Manhattan Plots for GWAS Data. 2021 [cited 2022 Feb 1]. Available from: https://CRAN.R-project.org/package=qqman

68. Goerg GM. LambertW: Probabilistic Models to Analyze and Gaussianize Heavy-Tailed, Skewed Data. 2020 [cited 2022 Feb 1]. Available from: https://CRAN.R-project.org/package=LambertW

69. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. 2019 Nov 14;20(1):238.

70. Copernicus Climate Change Service. E-OBS daily gridded meteorological data for Europe from 1950 to present derived from in-situ observations. ECMWF; 2020 [cited 2022 Feb 9]. Available from: https://cds.climate.copernicus.eu/doi/10.24381/cds.151d3ec6

71. Pierce D. ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files. 2021 [cited 2022 Feb 1]. Available from: https://CRAN.R-project.org/package=ncdf4

72. Hijmans RJ, Phillips S, Elith JL and J. dismo: Species Distribution Modeling. 2021 [cited 2022 Feb 1]. Available from: https://CRAN.R-project.org/package=dismo

73. Ziyatdinov A, Vázquez-Santiago M, Brunel H, Martinez-Perez A, Aschard H, Soria JM. lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. BMC Bioinformatics. 2018 Feb 27;19(1):68.

74. Jühling F, Kretzmer H, Bernhart SH, Otto C, Stadler PF, Hoffmann S. metilene: fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. Genome Res. 2016 Feb 1;26(2):256–62.

75. Kruijer W, Kooke with a contribution from IW (the internal function pin) C data collected by PF and R. heritability: Marker-Based Estimation of Heritability Using Individual Plant or Plot Data. 2019 [cited 2022 Feb 1]. Available from: https://CRAN.R-project.org/package=heritability