MOLECULAR ECOLOGY

Molecular Ecology (2016) **25**, 1823–1837 doi: 10.1111/mec.13573

EPIGENETIC STUDIES IN ECOLOGY AND EVOLUTION

The multivariate association between genomewide DNA methylation and climate across the range of *Arabidopsis thaliana*

THOMAS E. KELLER,* JESSE R. LASKY†‡ and SOOJIN V. YI*

*School of Biology, Institute for Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, GA 30332, USA, †Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY 10027, USA, ‡Department of Biology, Pennsylvania State University, University Park, PA 16802, USA

Abstract

Epigenetic changes can occur due to extracellular environmental conditions. Consequently, epigenetic mechanisms can play an intermediate role to translate environmental signals to intracellular changes. Such a role might be particularly important in plants, which often show strong local adaptation and have the potential for heritable epigenetic states. However, little is currently known about the role of epigenetic variation in the ecological mechanisms of adaptation. Here, we used multivariate redundancy analyses to examine genomewide associations between DNA methylation polymorphisms and climate variation in two independent panels of Arabidopsis accessions, including 122 Eurasian accessions as well as in a regional panel of 148 accessions in Sweden. At the single-nucleotide methylation level, climate and space (geographic spatial structure) explain small yet significant amount of variation in both panels. On the other hand, when viewed in a context of genomic clusters of methylated and unmethylated cytosines, climate and space variables explain much greater amounts of variation in DNA methylation than those explained by variation at the single-nucleotide level. We found that the single-nucleotide methylation polymorphisms with the strongest associations with climate were enriched in transposable elements and in potentially RNA-directed methylation contexts. When viewed in the context of genomic clusters, variation of DNA methylation at different sequence contexts exhibit distinctive segregation along different axes of variation in the redundancy analyses. Genomewide methylation showed much stronger associations with climate within the regional panel (Sweden) compared to the global (Eurasia). Together, these findings indicate that genetic and epigenetic variation across the genome may play a role in response to climate conditions and local adaptation.

Keywords: adaptation, bioinfomatics/phyloinfomatics, ecological genetics, population ecology

Received 27 August 2015; revision received 30 November 2015; accepted 17 December 2015

Introduction

Epigenetic mechanisms determine how genomic DNA is packaged and accessed in cellular environment, thus affecting patterns of cell-type-specific gene expression. It is increasingly recognized that substantial amounts of epigenetic variation exist in natural populations (Heyn *et al.* 2013; Moen *et al.* 2013; Schmitz *et al.* 2013). Whether such epigenetic variability can be transmitted across generations is currently under debate. In plants, there are several well-characterized examples of heritable transmission of epigenetic states across multiple generations (Kalisz & Purugganan 2004; Johannes *et al.* 2009; Cortijo *et al.* 2014; Jiang *et al.* 2014). Heritable epigenetic variation may provide a mechanism for adaptive evolution in response to natural selection. Furthermore, epigenetic variation determined by genetic

Correspondence: Thomas E. Keller, Fax: 404 894 0519; E-mail: thomas.e.keller@gmail.com

variation may cause expression polymorphisms involved in adaptation. However, little is currently known about the role of epigenetic variation in the ecological mechanisms of adaptation. Here we study this topic by examining the association between genomewide DNA methylation polymorphism and local climates of natural ecotypes of *Arabidopsis thaliana.*

DNA methylation is one type of epigenetic modification whose state can be transmitted to offspring generations (Schmitz *et al.* 2011; Calarco *et al.* 2012). DNA methylation varies across the genome as well as between individuals in many taxa (Zilberman *et al.* 2006; Lister *et al.* 2008; Becker *et al.* 2011; Schmitz *et al.* 2011, 2013; Dubin *et al.* 2015). In particular, the recently developed whole-genome bisulphite-sequencing methodologies allow researchers to characterize DNA methylation at the scale of single nucleotides. Studies suggest that many single-nucleotide methylation polymorphisms (SMPs) and differentially methylated regions (DMRs, regions of DNA that exhibit multilocus polymorphism in levels of methylation) may have underlying genetic determinants (Schubeler 2015). At the same time, others appear entirely 'epigenetic', that is their variation is largely independent of underlying genetic mechanisms (Schmitz *et al.* 2013). Indeed, it has been long postulated that epigenetic variation may be caused by both genetic and environmental factors (Liu *et al.* 2008).

Local adaptation to environment is widespread in nature, although much remains unknown about the genomic basis of local adaptation. For example, the spatial scale of local adaptation is poorly known, *that is* whether locally adaptive variants are geographically widespread or narrowly distributed (Fournier-Level *et al.* 2011; Lasky *et al.* 2015). Additionally, the role of genomewide regulatory evolution in local adaptation is little understood (Fraser 2013; Lasky *et al.* 2014). Local adaptation is likely to be particularly common in plants, as their migration is restricted to the seed stage (Herman *et al.* 2014), limiting their ability to avoid local environmental stressors. Epigenetic variation may be important in plant local adaptation, because epigenetic mechanisms could rapidly translate regional environmental conditions into potentially heritable changes at the cellular level [Liu (2013), but see Hagmann *et al.* (2015)]. Consequently, an emerging question is whether and to what extent epigenetic variation is involved in local adaptation to environmental conditions such as climate. Using genomewide SMP data, we may begin to assess how epigenomes interact with organismal environments, which is critical in understanding to what extent ecological factors contribute to epigenetic diversity and ultimately evolution.

We investigate whether climate can explain epigenetic variation in *Arabidopsis thaliana* by utilizing two recently generated whole-genome methylation panels of natural ecotypes sampled across Eurasia (Schmitz *et al.* 2013) and separately within Sweden (Long *et al.* 2013; Dubin *et al.* 2015). As these data also include an analysis of genomewide SNPs in the same accessions, we can compare the relative association of climate with epigenetic vs. genetic variation. We found that climate variables were statistically associated with a comparable amount of epigenetic and genetic variation in these data sets. The largest axis of methylation-climate association was strongly linked to CHH methylated sites and concentrated in transposable elements. In contrast, CG methylation associated with climate localized to genic regions. Previous analyses of *Arabidopsis* have suggested that local adaptation to climate may explain a substantial portion of observed spatial genetic structure (Hancock *et al.* 2011; Lasky *et al.* 2012; Long *et al.* 2013). Our results suggest that a substantial portion of natural epigenetic variation is also associated with climate and may be involved in local adaptation to climate.

Materials and methods

Genome and epigenome data collection

We used two main epigenomic data sets of Arabidopsis accessions in Europe. The first data set, hereafter referred to as the 'Eurasian panel', is a collection of preprocessed genomewide methylation maps for 140 *Arabidopsis thaliana* accessions from Schmitz *et al.* (2013) (NCBI GEO Accession no. GEO43857). The majority of these methylation maps were generated from leaf tissues, with a minority being from inflorescence tissues. Whole-genome variant calls (compared to the Col-0 reference strain) from the same study were obtained from the NCBI SRA database (Accession no. SRA012474). We constrained our analyses to 122 Eurasian accessions with well-defined site locations (Anastasio *et al.* 2011) so that climate variables could be unequivocally assigned to each accession. SMP analyses were restrained to sites that had at least one site differentially methylated among accessions after the binomial correction following Schmitz *et al.* (2013). We additionally excluded sites that lacked methylation status for any accession. This first filtering procedure changed the original methylation composition of SMPs from 14% CG, 16% CHG, and 70% CHH to 54% CG, 17% CHG and 29% CHH, indicating that many differentially methylated sites are found in CG context.

In the original Eurasian panel data set, approximately 3.5 million SNPs were segregating. To make computational analyses feasible, we first removed rare SNPs by

filtering out sites with a minor allele frequency <5%. To further reduce the number of SNPs and to avoid linked sites, we then used PLINK (Purcell *et al.* 2007) to remove sites with high linkage disequilibrium with other sites. We used a sliding window of 100 sites with a maximum pairwise correlation of 0.5 and a step size of five sites per sliding window, similar to previous studies (e.g. O'Connor *et al.* 2015). We then used the same filtering procedures (removing rare alleles and linked sites) for SMPs to facilitate fair comparisons among data sets. After these filtering steps, the SMP data set had 182 090 sites (62% CG, 6% CHG, 32% CHH) while the SNP data set had 328 795 sites.

The second epigenomic data set, later denoted as the 'Swedish panel', consists of whole-genome methylation maps for 148 accessions from across Sweden (Dubin *et al.* 2015). In this study, two groups of accessions were raised at two different temperatures. We selected accessions grown at 10 °C to utilize the larger sample size compared to the other and also because it more closely reflects the growing season conditions in this region (Lasky *et al.* 2012). The SMP compositions of unfiltered data set were 9% CG, 16% CHG and 75% CHH, suggesting some difference in the standing level of CG methylation variation between populations. After removing sites with static methylation or missing data, the composition was 20% CG, 19% CHG and 61% CHH, indicating that CG sites tend to be differentially methylated between individuals, similarly to the observation from the Eurasian panel. After performing PLINK pruning as in the Eurasian panel, the SMP data set included 162 544 sites (26% CG, 10% CHG, 64% CHH). Thus, in both panels, DNA methylation at CG sites appears to be highly variable across individuals, found in appreciable frequencies (>5%) and unlinked from other variation at CG sites. On the other hand, DNA methylation variation at CHG and CHH sites was reduced following the filtering steps. In addition, there were significant differences at the initial level of standing variation of DNA methylation at CG sites. The causes of these patterns are currently unknown, and with more population epigenomic data sets in the future, we can address the robustness and significance of such differences.

To directly compare variance explained by epigenetic and genetic data in this panel, we also analysed a SMP data set restricted to the 94 accessions from (Long *et al.* 2013) with accompanying genomic data from (Dubin *et al.* 2015). In this restricted data set of 94 accessions, the number of SMPs was 162 609 variable sites, while the corresponding SNP data set had 304 720 sites. The final list of accessions used is listed in Table S1 (Supporting information).

Controlling for potential batch effects is an important aspect of data analyses. The whole-genome bisulphite-sequencing data used in this study were generated on the next-generation sequencing platform, which is known to be robust against batch effects, unlike microarray-based platforms (Marioni *et al.* 2008). Consequently, no classifier associated with potential batch is provided with these data. Nevertheless, these data included false-positive controls within each experiment. For the Eurasian panel (Schmitz *et al.* 2013), the average nonconversion rate was 0.2%, while in the Swedish panel (Dubin *et al.* 2015) it was 0.41%. Additionally, Dubin *et al.* (2015) sequenced 11 biological replicates and reported low variation across methylation contexts.

Identification of DMRs

We identified DMRs across the accessions following Schmitz *et al.* (2013) prior to pruning. Briefly, we conducted a sliding window analysis using a 100-base window for DMRs comprised of CG, CHG and CHH methylation sites (where H is any base other than G), defined here as C-DMRs, and a 300 base window for DMRs consisting of just CG sites (CG-DMRs). These DMR types are often considered separately in genomewide DNA methylation analyses due to their distinctive nature with respect to genomic locations and contexts. For example, it was previously shown that C-DMRs mainly localize to the repeat-rich centromeres and CG-DMRs to the more gene-rich interior of chromosome arms (Schmitz *et al.* 2013). A preliminary DMR was called if at least 10 sites containing single methylation polymorphisms (SMPs) within a window had methylated and unmethylated accessions, as determined by the DNA methylation binomial test [a site was called methylated if it harboured a greater number of methylated reads than expected under the binomial distribution where the probability of failure was the observed error rate due to incomplete bisulphite conversion (Lister *et al.* 2008)]. The fractional methylation at each site was then used to calculate whether accessions differed in methylation using a Kruskal–Wallis test; accessions with <5 methylation values in the preliminary DMR were excluded from this calculation to avoid errors due to small sample size (Kruskal & Wallis 1952). Preliminary DMRs were joined if they were within 50 bases of one another; the *P*-value of the joined DMRs was calculated using Fisher's method with the two previous *P*-values. We then applied a 1% false-discovery rate (FDR) adjustment to these DMRs (Benjamini & Hochberg 1995). Finally, CG-DMRs overlapping with C-DMRs were excluded.

Climate data collection

We used publically available global climate data sets to characterize the home environment of each accession [locations taken from Dubin *et al.* (2015), Schmitz *et al.* (2013)]. We primarily used data from WORLDCLIM, a database with 30-arcsecond resolution (approximately 1 square kilometre) (Hijmans *et al.* 2005). WORLDCLIM is a global weather map of average conditions from the years 1950–2000, interpolated using data accumulated from weather stations around the world. Monthly averages for precipitation as well as minimum, average and maximum temperature were calculated for temperature. WORLDCLIM also provides 19 additional climate variables of biological importance derived from monthly conditions, as well as altitude (Hijmans *et al.* 2005). We also included two additional variables calculated using WORLDCLIM data, annual total potential evapo-transpiration (PET) and aridity index (annual precipitation/PET) (Zomer *et al.* 2008).

Following Lasky *et al.* (2012), we calculated the growing season for each accession using the WORLDCLIM precipitation and temperature data (Walter & Leith 1960). For each location, the growing season was defined as the set of months with an average precipitation (mm) that was greater than or equal to twice the average temperature (Celsius), with the average temperature being at least 4 °C. From these estimates, we calculated growing season length as well as mean precipitation, total precipitation, mean temperature, minimum temperature, maximum temperature and the coefficient of variation for precipitation during the growing season.

As monthly temperature and precipitation estimates are highly correlated, we considered these values by quarter (January, April, July, October). This filtering reduced the number of climate variables well below the number of accessions analysed. We attempted several further attempts at reducing model complexity, but the overall significance remained similar.

Estimating spatial structure of Arabidopsis accessions

Genetic variation among accessions may be associated with geographical distance, which we modelled using principle components of neighbourhood matrices (PCNM) (Borcard & Legendre 2002; Manel *et al.* 2010; Lasky *et al.* 2012). PCNM are variables estimating different axes of spatial relationships among accessions, from very large to very small scales, and allow for modelling of nonstationary isolation by distance. First, a distance matrix between locations was calculated using the Vincenty Ellipsoid formula from the R 'geosphere' package. Second, a minimum-spanning tree (MST) between locations was constructed using these distances. Distances between locations that were longer than the maximum distance between points in the MST were truncated to four times the maximum MST distance. Truncating distances in this fashion prevents the eigenvector calculations from being dominated by long-distance structure (Borcard & Legendre 2002).The principle coordinates of neighbourhood matrix (PCNM) was calculated using this threshold with the 'pcnm' function implemented in the R package 'VEGAN' (Oksanen *et al.* 2015). After removing PCNM axes with negative eigenvalues, there were 61 variables used in the Eurasian panel and 47 in the Swedish panel.

Statistical estimation of genetic and climate associations

We estimated the amount of methylation (SMP or DMR) and genetic (SNP) variation that could be explained by climate and geography using redundancy analysis (RDA) and variance partitioning, also implemented in 'VEGAN' (Van Den Wollenberg 1977; Peres-Neto *et al.* 2006; Oksanen *et al.* 2015). RDA is a regression technique that models the relationship between multivariate predictors (here climate and spatial variables) and multivariate responses (here SMP, DMR or SNP matrices). RDA can also partial out sets of explanatory variables, which allowed us to control for spatial structure while modelling associations with environment as an effort to remove effects of geographic population structure. RDA maximizes the amount of variance in a linear combination of response variables explained by linear combinations of explanatory variables. RDA is similar to principle components analysis in that it is an eigenanalysis that produces orthogonal axes of variation, referred to as canonical axes. We calculated the portion of variance explained by climate and by spatial variables, and the portion explained by climate collinear with space.

In this study, we present results of analyses including all 42 climate variables. Considering such similarities, and unclear associations between different climate and space variables, we decided to include all variables without specific a priori exclusion of specific variables. We also note that since RDA explicitly models similar correlations between multiple predictors and the response variables, predictors that are correlated in such a manner load similarly on the resulting canonical axes, as in PCA or other eigenanalyses.

Assessing significance of associations

We assessed the significance of the estimated variance explained in response variables using permutations to generate a null distribution of variance explained. We

calculated empirical *P*-values by comparing the observed variation explained to that explained by null matrixes randomized by free permutation with the VEGAN function 'anova.cca' (Oksanen *et al.* 2015). The *P*-value was then calculated as the proportion of 1000 random matrices that had a higher variance explained than the observed data.

To assess whether certain methylation contexts or genomic regions were enriched for climate-associated variation, we generated null distributions by circularizing the genome and then permuting the methylation context (CG, CHH and CHG) or genomic context with respect to gene annotation (e.g. intergenic, promoter, gene body or transposable element) to a random position in the data set. This procedure preserves the linkage disequilibrium and order of the original data. Empirical *P*-values were calculated as in the previous analysis. The gene annotations were based off the TAIR10 gene and transposable element locations (Lamesch *et al.* 2011). Promoter regions were defined as 1500 bases upstream and 500 bps downstream of a transcription start site (Zeng *et al.* 2012).

In our analyses, we observed that methylation contexts differed dramatically in how much genomewide variation was explained by the first canonical climate axis. This pattern suggests the dimensionality of genomewide variation differs among methylation contexts. To assess the significance of these patterns, we compared the variance explained by the first canonical RDA axis in null permutations to the variance explained by the first canonical RDA axis in the observed data for each methylation context. We used a one-tailed test to estimate a *P*-value.

We calculated the proportion of variation explained by a specific climate or spatial variable in an RDA as P_x, following (Lasky *et al.* 2012). P_x was calculated as the weighted sum of absolute correlations across all canonical axes. Each absolute correlation was weighted by the proportion of variation explained by a canonical axis (e.g. the eigenvalue).

Gene ontology enrichment among climate-associated loci

We tested whether climate-associated SMPs and DMRs overlapped with genes that had specific functions by performing gene ontology (GO) enrichment analyses using AGRIGO (Du *et al.* 2010), using a hypergeometric test and FDR correction under dependency (Benjamini & Yekutieli 2001). For the top 1% climate-associated SMPs, C-DMRs and CG-DMRs, we assembled a list of genes with overlapping promoter or gene body regions. This list of genes was compared against the Slim TAIR10 set of genes for *A. thaliana*.

Analyses of gene expression

To understand the potential relationship between climate-associated methylation and gene expression, we analysed RNA-seq data of the same accessions. RNA-seq gene expression values were obtained from (Schmitz *et al.* 2013) and Dubin *et al.* (2015). We used the final estimates from both studies (FPKM and RPKM, respectively). The numbers of accessions with matching RNA-seq data sets are 107 (of 122) in the Eurasian panel and 135 (of 148) in the Swedish panel. We restricted this analysis to DMRs that uniquely mapped to a single promoter or gene body region, and considered each region separately. We then calculated the average absolute correlation between gene expression and DMR methylation for each data set.

Associations between Gene × Environment effects and climate

We also examined the correlation between methylation and expression for a subset of genes that were identified to have strong Genetic × Environment (G × E) effects in drought and cold conditions as measured by gene expression differences (Des Marais *et al.* 2012; Hannah *et al.* 2006; Lasky *et al.* 2014). For each data set, we calculated the average squared loading for all DMRs or SMPs that were within 10 KB of G × E genes. We then compared the observed average with the 1000 bootstrap replicates containing the same number of random loci as the observed data set.

Results

DNA methylation is associated with climate and geography at the single-nucleotide level

We began using RDA to estimate how much genomewide variation in single-nucleotide polymorphisms (SNPs) and single methylation polymorphisms (SMPs) could be explained by climate and spatial variables in the two panels. The results from the two panels revealed intriguing similarities and differences (Table 1). First, in the full model, climate and space explained 1% (in Swedish panel, derived from Dubin *et al.* 2015) and 7.5% [in Eurasian panel, derived from Schmitz *et al.* (2013)] of total genetic variation (SNP), respectively. The amount of genetic variation explained by climate and space is much lower than what was previously observed in a larger panel of accessions (Lasky *et al.* 2012). Nevertheless, the effects of climate and space are significant in this model.

Interestingly, the amount of total genomewide epigenetic variation (SMP) explained by climate and space was of a similar order to that for genetic variation. For

Table 1 Single-nucleotide methylation polymorphism (SMP) and SNP variation explained by redundancy analysis (RDA). Adjusted R^2 and significance are reported for both a full model (climate + space) and a model of climate independent of space (climate | space). These models were implemented with the VEGAN R package. Significance was calculated as the proportion of 1000 permuted null data sets that exceed the observed variance explained. The Eurasian data set was comprised of 122 SMP and SNP samples while the Swedish data set had 148 SMP samples and 94 SNP samples

		Full Model (Climate + Space)		Space-adjusted Model	
Data sets	Response variables	Adj. R^2	*P*-value	Adj. R^2	*P*-value
Eurasian panel	SMP	0.025	0.023	0.010	0.335
Schmitz *et al.* (2013)	SNP	0.074	0.001	0.034	0.115
	CG SMP	0.033	0.001	0.021	0.214
	CHG SMP	0.031	0.058	0.003	0.432
	CHH SMP	0.011	0.393	0.000	0.589
	Genebody	0.031	0.001	0.019	0.244
	Intergenic	0.008	0.385	0.000	0.600
	Promoter	0.024	0.039	0.009	0.378
	Transposon	0.017	0.348	0.000	0.583
Swedish panel	SMP	0.076	0.001	0.046	0.001
Dubin *et al.* (2015)	SNP (94 samples)	0.010	0.253	0.014	0.276
Long *et al.* (2013)	SMP (94 samples)	0.050	0.001	0.030	0.075
	CG SMP	0.117	0.001	0.067	0.001
	CHG SMP	0.103	0.001	0.063	0.001
	CHH SMP	0.050	0.001	0.034	0.002
	Genebody	0.097	0.001	0.058	0.001
	Intergenic	0.094	0.001	0.056	0.001
	Promoter	0.116	0.001	0.067	0.001
	Transposon	0.045	0.001	0.031	0.001

example, in the Eurasian panel, climate and space explained 2.5% of total SMP variation, while in the Swedish panel it was 5% (Table 1). When we accounted for spatial structure by conditioning the model on PCNM variables, the amount of variation explained by climate decreased two- to threefolds in both data sets. An exception was CHG SMPs in the Eurasian panel, for which the portion explained by climate declined an order of magnitude (3.1–0.32%) in the space-adjusted model compared to the full model.

Aside from these similarities, the explanatory power of climate variables varied greatly between the two data sets. In the Eurasian panel, all climate variables became nonsignificant ($P > 0.05$) when we partialled out effects of the spatial variables. In contrast, the comparable models using the Swedish panel generally remained highly significant (Table 1). The relative explanatory power of individual climate variables (P_x, see Methods) also varied between data sets (Table 2). The climate variables with the highest P_x were generally related to temperature in the Eurasian panel and precipitation in the Swedish panel (Table 2). The sample sites and gradient of the strongest climate variable for each panel are shown in Fig. 1.

The number of climate or spatial variables does not strongly affect explanatory power

We examined reduced data sets to explore whether we could reduce the complexity of the data. However, the variance explained by the reduced data sets was generally similar to the variance explained by the full data set. For example, the adjusted R^2 in the Eurasian data set explained by climate and space was 0.025 with all 42 climate variables, remaining 0.025 when only considering 36 WORLDCLIM variables, and finally declining slightly to 0.020 when only considering altitude and the 19 nonmonthly WORLDCLIM variables.

We also experimented with the effect of reduced spatial variables. The explanatory power of the statistical models remained similar in the Eurasian panel; the adjusted R^2 of the full model changed by only 0.001 when reducing the PCNM variables from 61 to the first 40, and by 0.002 when they were reduced to the first 20. Similarly, in the Swedish panel, the adjusted R^2 changed by 0.001 when reducing the PCNM variables from 47 to 35, and 0.004 when there were 25 PCNM variables.

Table 2 Climate variables and the percentage of single methylation polymorphism (SMP) variation they explain in redundancy analysis (RDA) full models. The percentage explained was calculated as the sum of the absolute correlations of a climate variable to each RDA axis and normalized by the total amount of variation explained by each axis (e.g. the eigenvalues). The top 10 climate variables are shown for each panel

Eurasian panel		Swedish panel	
Climate variable	Per cent of SMP variation explained	Climate variable	Per cent of SMP variation explained
Isothermality	7.39	Annual prec.	6.08
Photosythetically active radiation (PAR) in summer quarter	7.18	Prec. of wettest quarter	6.04
Annual mean temperature	7.09	Prec. of coldest quarter	6.01
Mean October temp.	7.06	Prec. of driest month	6.01
Min. April temp.	7.05	Prec. of driest quarter	6.01
Prec. seasonality	7.00	Prec. of warmest quarter	6.00
Mean April temp.	6.99	Isothermality	5.98
Max October temp.	6.99	January prec.	5.98
Prec. Of driest mo.	6.98	July prec.	5.97
Max January temp.	6.94	October prec.	5.96
Min. October temp.	6.94	Photosythetically Active Radiation (PAR) in summer quarter	5.96
Mean temp. of coldest quarter	6.93	Aridity	5.89
Mean January temp.	6.90	April prec.	5.84
Max April temp.	6.89	Prec. of warmest month	5.74
Growing season length	6.89	Prec. seasonality	5.67
Prec. of warmest quarter	6.85	Mean temp. of wettest quarter	5.49

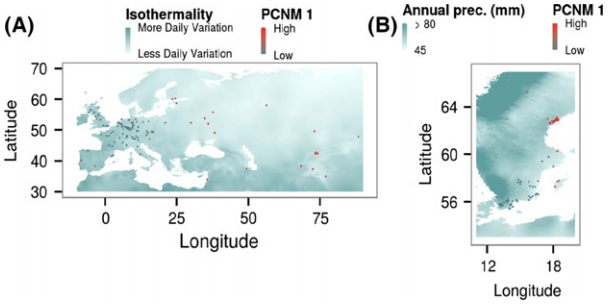

Fig. 1 Distribution of 122 Eurasian accessions and 148 Swedish accessions used in separate redundancy analyses of methylation variation [redundancy analysis (RDA)]. (a) Eurasian accessions from Schmitz *et al.* (2013) are coloured by the first principle components of neighbourhood matrices (PCNM) variable, which describes spatial structure between samples. These samples are shown with isothermality (defined as the average daily variation in temperature relative to the seasonal variation in temperature) which explained the most single-nucleotide methylation polymorphism (SMP) variation in an RDA analysis. (b) Swedish samples from Dubin *et al.* (2015) coloured by the first PCNM variable as in (a). These samples are shown with annual precipitation which explained the most SMP variation in an RDA analysis.

Climate and space explain more DMR variation than SMPs independently

We then examined how DMRs vary with climate and space using RDA (Table 3). In the Eurasian panel, C-DMRs had an average length of 212 (±1.4 [SE]) bases and covered 6.1% of the genome. A substantial amount of variation (6.5%) at C-DMRs was explained by climate and space in this panel ($P = 0.002$). DMRs in a CG context, CG-DMRs, were on average longer than C-DMRs (292 (±1.0 [SE]) bases) and occupied 5.9% of the genome. Similarly, a substantial amount of variation of DNA methylation (4.7%) at CG-DMRs was explained by climate and spatial variables ($P = 0.01$).

In the Swedish panel, C-DMRs tend to be longer than observed in the Eurasian panel (average of 428 (±2.1 [SE]) bps compared to 212 bps (±1.8 [SE]) in the latter panel) and also occupy a much greater portion (22%) of the genome. In contrast, CG-DMRs in the Swedish panel are of similar lengths to those in the Eurasian panel (average length of 287 (±2.1 [SE] bps) but occupy much shorter regions of the genome (0.4%). This difference may be partially due to the less frequent polymorphic CG methylation relative to the total number of SMPs in the Swedish panel (2.02 vs. 2.97 million sites, 13% vs. 31% of total SMPs in the Swedish and the Eurasian panels, respectively). We found that climate and space explain a substantial amount of variation of DNA methylation in DMRs (16% and 18% for C-DMR and CG-DMR, respectively).

The increase in variance explained by DMRs compared to individual SMPs in both panels was striking, possibly due to the difference in the numbers of

Table 3 Differentially methylated region (DMR) methylation variation explained by redundancy analysis (RDA) in Eurasian and Swedish panels. Significance was assessed similarly to Table 1

		Full Model (Climate + Space)		Space-adjusted Model	
Data sets	Response variables	Adj. R^2	*P*-value	Adj. R^2	*P*-value
Eurasian panel	C-DMRs	0.07	0.002	0.03	0.182
Schmitz *et al.* (2013)	CG-DMRs	0.05	0.010	0.03	0.170
Swedish panel	C-DMRs	0.16	0.001	0.09	0.001
Dubin *et al.* (2015)	CG-DMRs	0.18	0.001	0.09	0.001

response variables analysed in the models (e.g. 182 090 SMP sites vs. only 38 443 C-DMRs in the Eurasian panel full model). To test the effect of sample size, we randomly selected 38 443 SMPs and ran the RDA, and repeated the procedure 1000 times. Results from this analysis were nearly identical to those from the full data set, indicating that the observed patterns were not caused by the difference in the sample sizes.

Epigenetic variance explained by climate depends on methylation and genomic context

After considering the global methylation association with climate, we subdivided SMPs depending on their methylation (CG, CHG, or CHH) or genomic context according to gene annotation (promoter, gene body, transposable element or intergenic SMP). We then assessed the explanatory power of the first canonical RDA axis by a permutation method (see Methods). The results from the first canonical RDA axis, which explains the greatest portion of variation in the response, illustrate different impacts of genomic and epigenomic contexts (Fig. 2). In both panels, CHH methylation and transposable element SMPs were strongly enriched on the first RDA axis.

Indeed, SMPs with the strongest loading on the first canonical RDA axis are overwhelmingly in the CHH context (Fig. 3A and D). In fact, 999 of the top 1000 climate-associated SMPs in the Eurasian panel and all 1000 in the Swedish panel were CHH polymorphisms. In contrast, the canonical axis 2 (recall all axes are orthogonal) is enriched for CG polymorphisms in both panels, whereas RDA axis 3 is most associated with CHG polymorphisms (Fig. 3). These consistent associations in methylation context also have functional implications; CHH methylation is mostly associated with transposable elements, while CG methylation is most prevalent in or near genes. We also compared how the proportion of variance explained by each axis varied by methylation context (Fig. S1, Supporting information). Notably, CHH methylation has a stronger loading on the first few axes of variation, while CHG and CG methylations have lower loadings on the first axes but higher loadings on subsequent axes. This pattern is most evident in the Eurasian panel (Fig. S1A, Supporting information).

CG-DMRs are enriched in terms relevant for local adaptation and potentially affect gene expression

We investigated whether genes overlapping with DMRs that are highly associated with climate and space variables (i.e. SMPs or DMRs in the 99th percentile of loadings on a given axis) were enriched for specific GO categories using AGRIGO (Du *et al.* 2010). We examined enrichment in the Eurasian panel using the full RDA model (climate + space). We found that no significant GO terms were identified in genes overlapping C-DMRs in either the promoter or gene body region. In contrast, genes overlapping with CG-DMRs in the full model were enriched for a variety of GO terms, including ones relating to response to abiotic stimulus, reproduction, development, and metabolism (Table 4). There was no enriched GO term for C-DMRs or CG-DMRs from the Swedish panel.

Gene expression correlates with climate-DMRs

One potential consequence of climate-associated DNA methylation difference is gene expression variation. We thus examined the expression variation of genes found in DMRs. Expression consequence of DNA methylation can vary according to the genomic context where DNA methylation occurs. For example, methylation of regulatory regions tends to suppress gene expression, while methylation of gene bodies is associated with increased level of gene expression. To assess the extent DNA methylation variation influences gene expression variation, we first computed the overall and absolute correlation between gene expression and DNA methylation of genes found in climate C-DMRs and CG-DMRs based

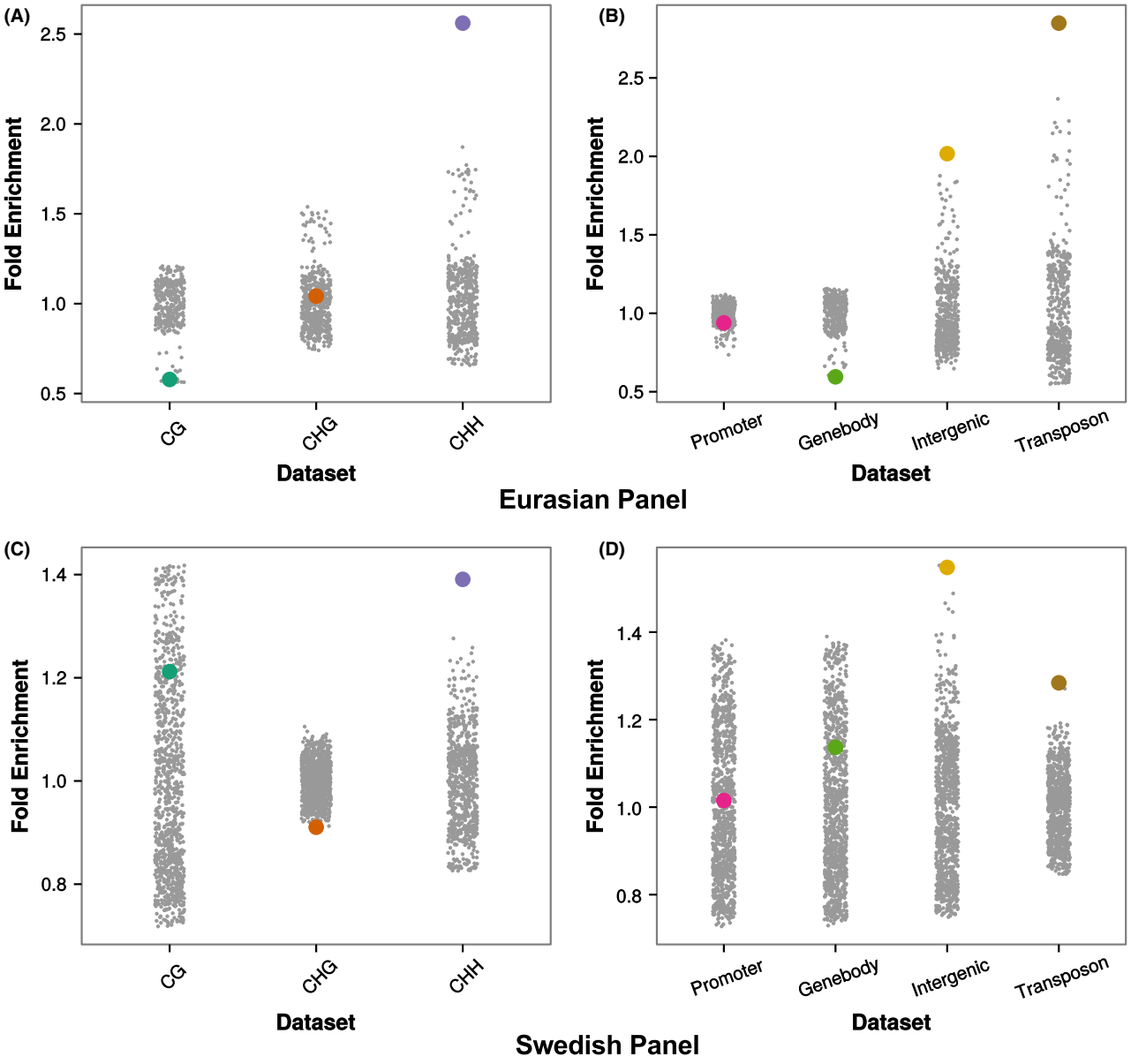

Fig. 2 Enrichment analysis of variance explained by different subsets of single-nucleotide methylation polymorphismS (SMPs) in the Eurasian and Swedish panels. The *y*-axis shows fold enrichment, which is the amount of SMP variation (measured by eigenvalue) explained by the first redundancy analysis (RDA) axis divided by the average amount explained in 1000 permuted data sets. Grey dots correspond to the null data sets for each category, while large dots are the observed estimates. (A, C) Fold enrichment of different methylation contexts in Eurasian and Swedish panels, respectively. (B, D) Fold enrichment of different genomic categories according to TAIR 10 gene and transposable element annotations for Eurasian and Swedish panels, respectively.

on whether they overlapped with a gene promoter or gene body region (Table 5). C-DMRs in both panels had a mean correlation near 0 while the absolute correlations were near 0.1 in both regions, indicating similar numbers of DMRs with positive and negative associations with expression. In contrast, CG-DMRs were on average positively associated with expression when located within gene bodies, as expected. The average absolute correlation between methylation and expression was generally stronger in climate-associated DMRs compared to the genomic average.

Genes with genotype by environment (G × E) interactions are enriched for climate-associated SMPs

Genes with variable responses to abiotic stress depending on genotype (G × E) may be an indication of natural selection towards local climate conditions. A recent

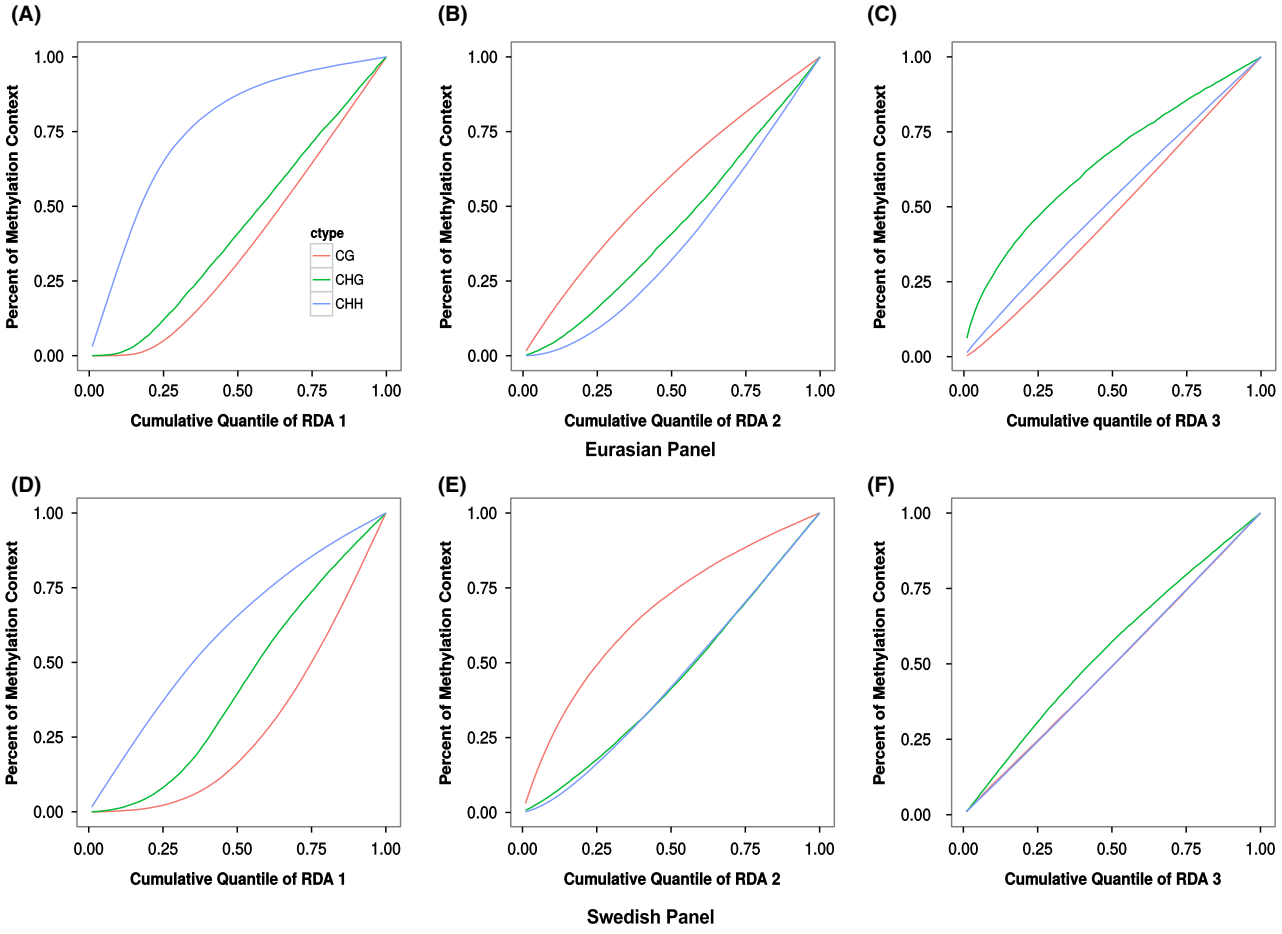

Fig. 3 Top redundancy analysis (RDA) axes are enriched by distinct methylation contexts. The *x*-axis represents quantiles of the data ordered by decreasing loading for each single-nucleotide methylation polymorphism (SMP). The *y*-axis represents the proportion of a given methylation context in each quantile. Thus, curves above a 1:1 line indicate an overabundance of that methylation context for a given quantile. (A–C) CHH, CG and CHG are enriched for RDA axes 1, 2 and 3 in the Eurasian panel. (D–F) CHH, CG and CHG are enriched for RDA axes 1, 2 and 3 in the Swedish panel, respectively.

study by Lasky *et al.* (2014) found that genes in *A. thaliana* with different expression responses to cold and drought conditions depending on genotype had more genetic polymorphisms in promoter regions and stronger associations with climate compared to stably-expressed genes. We analysed these same genes to determine whether climate associations with methylation loci were stronger near these genes compared to other genes given that methylation can also have a strong effect on expression. Specifically, we compared the model loadings of methylation loci within 10 KB of G × E genes to random sets of loci (Table 6). While C-DMRs and CG-DMRs did not have stronger climate associations near G × E genes, the second RDA axis of the SMP model (primarily CG methylation) had a stronger average loading than all random sets of genes ($P < 0.001$) in both cold and drought G × E gene sets as well as both Eurasian and Swedish panels. This analysis, along with the previous observation that CG SMPs overall association with climate, suggests that methylation as well as genotype may be relevant for local climate adaptation in gene expression.

Discussion

The recent whole-genome methylation maps of *A. thaliana* (Schmitz *et al.* 2011; Dubin *et al.* 2015) provide rich opportunities to investigate the associations between ecological variability and epigenetic variability. However, to date little is known about the role of DNA methylation in adaptation to environment (but see Dubin *et al.* 2015; Platt *et al.* 2015). As a first step to understanding the role of DNA methylation variation in local adaptation, here we examined how genetic and epigenetic variations are associated with climate and spatial variables in two large data sets.

Table 4 Gene ontology enrichment analysis on Eurasian panel associates CG-DMRs with genes. Fold enrichment is the proportion of genes in the outlier data set associated with the GO term divided by the proportion of genes in the background associated with the GO term

GO Category	Fold enrichment	False-discovery rate
Cell cycle	2.53	0.003
Post-embryonic development	1.96	0.004
Protein modification process	1.80	0.004
Regulation of gene expression, epigenetic	2.75	0.007
Developmental process	1.64	0.008
Multicellular organismal development	1.66	0.008
Multicellular organismal process	1.63	0.010
Cellular component organization	1.66	0.010
Biological regulation	1.53	0.010
Macromolecule modification	1.62	0.016
Protein metabolic process	1.53	0.016
Cellular developmental process	1.95	0.017
Regulation of biological process	1.51	0.021
Cellular protein metabolic process	1.53	0.021
Catabolic process	1.72	0.021
Cell differentiation	2.01	0.025
Reproduction	1.65	0.033
Response to abiotic stimulus	1.63	0.033
Regulation of biological quality	1.78	0.041
Anatomical structure development	1.54	0.041
Reproductive process	1.62	0.046
Homoeostatic process	2.47	0.048
Macromolecule metabolic process	1.35	0.048
Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	1.96	0.042
Hydrolase activity, acting on acid anhydrides	1.93	0.042
Pyrophosphatase activity	1.96	0.042
Receptor activity	4.97	0.042
Nucleoside-triphosphatase activity	1.94	0.042

Our analyses illustrate that spatial gradients in temperature and precipitation are associated with a significant portion of epigenomic variation across the native range of *Arabidopsis* (in the Eurasian panel), as well as at the regional scale (in the Swedish panel). The relative associations of climate variables with genetic vs. epigenetic variation appear to depend on the geographic scale considered as well as sample size. For example, an earlier study (Lasky *et al.* 2012) found that climate and geography explained substantial fraction of genetic diversity (22%) in a panel of *Arabidopsis* accessions across Eurasia, while in our Eurasian panel the amount of SNP variation explained was only 32% of the previous value. This discrepancy may be due to the difference in the number of genotypes used ($N = 1003$ in the previous study vs. $N = 122$ in the current study, respectively). We experimented with different reduced combinations of climate and spatial variables and found it had little effect on the amount of diversity explained, suggesting that differences in that part of the statistical model is unlikely to account for the difference in explanatory power.

Here, we found stronger climate associations for Swedish SMPs compared to SMPs in the Eurasian panel, even after partialling out spatial variables. This finding may suggest that genetic and epigenetic mechanisms of local adaptation may involve a larger portion of the genome in Sweden. Additionally, climate is likely to be highly spatially structured at the large scale (Eurasia) but less so at the regional scale (Sweden). Furthermore, mechanisms of local adaptation may be restricted geographically (Fournier-Level *et al.* 2011) such that global models obscure patterns occurring within regions (Lasky *et al.* 2015). It is also notable that similarly disproportionately stronger climate associations were found in the Scandinavian SNPs compared to populations in other regions (Lasky *et al.* 2012).

Table 5 Correlations between DMRs and gene expression. The average and absolute average correlation between methylation and expression for each climate-associated DMR that overlapped a gene body or promoter region compared to the genomic background correlation. We only considered DMRs that strictly overlapped a single promoter or a single gene body region. NA is listed where there were no genes to analyse

Panel	DMR	Region	Climate correlation	Absolute climate correlation	Genomic correlation	Absolute genomic correlation
Eurasia	C-DMR	Promoter	0.000	0.111	−0.002	0.092
Eurasia	C-DMR	Genebody	−0.071	0.099	0.012	0.090
Eurasia	CG-DMR	Promoter	0.055	0.055	0.003	0.088
Eurasia	CG-DMR	Genebody	0.050	0.103	0.020	0.088
Sweden	C-DMR	Promoter	−0.001	0.085	0.004	0.073
Sweden	C-DMR	Genebody	−0.041	0.126	0.017	0.077
Sweden	CG-DMR	Promoter	NA	NA	−0.041	0.091
Sweden	CG-DMR	Genebody	0.044	0.071	0.029	0.078

Table 6 Climate associations are stronger in CG single-nucleotide methylation polymorphisms (SMPs) near genes with G × E effects. With the average model loading for loci within 10 KB of genes with G × E interactions in either drought or cold conditions were compared to the average loading of 1000 bootstrap replicates with the same number of loci as the observed data. Enrichment was calculated as the observed average loading divided by the mean loading across replicates. *P*-values were calculated as the number of replicates with a higher average loading than the observed data divided by 1000

Dataset	Drought enrichment	Drought *P*-value	Cold enrichment	Cold *P*-value
Swedish C-DMRs	0.988	0.547	0.953	0.808
Swedish CG-DMRs	0.736	0.74	1.046	0.372
Swedish SMPs RDA 1	0.541	1	0.660	1
Swedish SMPs RDA 2	1.523	0	1.267	0
Swedish SMPs RDA 3	0.961	0.944	0.929	0.999
Eurasian C-DMRs	0.944	0.8	1.032	0.265
Eurasian CG-DMRs	0.962	0.864	0.982	0.712
Eurasian SMPs RDA 1	0.736	1	0.965	0.972
Eurasian SMPs RDA 2	1.154	0	1.117	0
Eurasian SMPs RDA 3	0.956	0.997	1.044	0.005

One common observation across the two panels is that variation in DNA methylation at clusters of differentially methylated sites between accessions, or DMRs, are more strongly associated with climate variables than individual methylation polymorphisms at the nucleotide level. In both panels, C-DMRs as well as CG-DMRs had approximately twice as much variance explained by climate compared to individual SMPs and this pattern could not be explained by differences in sample sizes of DMRs and SMPs. Instead, combined epigenetic variation of adjacent genomic positions (e.g. DMR) may be of greater significance than variation at individual CpG sites. Recent genomewide analyses often identify regions of adjacent cytosines whose coordinated epigenetic variation is critical in biological processes (e.g. Elliott *et al.* 2015). If the combined influence of polymorphism at multiple loci is more functionally important, it follows that computational approaches (e.g. association studies) should model variation at multiple loci simultaneously. For example, many current association studies and genome scans focus on variation at single nucleotides independent of other loci (but see Segura *et al.* 2012).

Most (>99%) of the top 1000 climate SMPs on the first RDA axis occurred at CHH sites in both data sets. These climate SMPs were primarily located within transposable elements. Their location suggests a possible functional link between TEs and adaptation to local climate conditions. Indeed, several recent studies also identified a strong association between temperature and TE methylation (Shen *et al.* 2014; Dubin *et al.* 2015). Specifically, temperature was associated with variation at CHH in TEs, which were caused by segregating SNPs within and surrounding CMT2, a TE-specific methyltransferase (Zemach *et al.* 2013; Stroud *et al.* 2014). Notably, some normally silenced TEs can be activated when Arabidopsis plants are exposed to prolonged conditions of extreme heat (Pecinka *et al.* 2010). Thus, our observation that the first RDA axis for variation of SMP methylation occurs at TE CHH sites are consistent with the role of CMT2 and TE methylation on adaptation to climate changes. However, it is important to note that the first axis of genomic variation in eigenanalyses (e.g. PCA, RDA) is often also strongly associated with population structure (Horton *et al.* 2012; Lasky *et al.* 2012; Nordborg *et al.* 2005). Consequently, the first RDA axis and the associated CHH variation may reflect underlying population structure, rather than a direct response to climate variables.

However, previous studies found significant associations between CHH methylation and temperature variables while accounting for population structure (Shen *et al.* 2014; Dubin *et al.* 2015). In our study, we observe the same pattern in the Swedish panel after accounting for spatial structure, a proxy for population structure (Sharbel *et al.* 2000; Platt *et al.* 2010). Experimental studies that manipulate methylation and test for methylation state-by environment interactions are warranted to further understand the role of methylation in local adaptation.

On the other hand, we found many climate- and space-associated CG-DMRs within genic regions. Additionally, CG SMPs overall had the strongest climate association overall and spread more evenly across RDA axes compared to CHH methylation, which had the strongest loading on the first axis. CG methylation was also found to more strongly differentiate oak populations compared to other contexts (Platt *et al.* 2015). In addition, Jiang *et al.* (2014) found that high salinity soil induced a ~45% increase in differentially methylated cytosine positions in the CG context compared to controls, with the great majority of these methylation

polymorphisms being inherited. The most climate- and space-associated CG-DMRs in the Eurasian panel were located in genes associated with several GO biological terms. Particularly notable terms were related to abiotic stimulus response, development and reproduction. Based on known ecophysiological mechanisms of moisture and temperature response, these GO terms are good candidates for mechanisms of local adaptation to climate. Nevertheless, enrichment of these GO terms is not a conclusive validation of our results (Pavlidis *et al*. 2012) and it is important to follow up with experimental validation of loci putatively involved in local adaptation.

Numerous recent studies have shown a genetic basis for local adaptation to various climate conditions (Hancock *et al*. 2011; Lasky *et al*. 2012; Long *et al*. 2013). Our reanalysis of two separate epigenomic data sets found broad similarities between the two panels with respect to associations between DNA methylation variation with climate, suggesting that these differences may contribute to local adaptation. A substantial amount of this methylation variation is likely to due underlying genetic differences (Dubin *et al*. 2015). However, epigenetic variation can be stably inherited in plants and thus may act as a further heritable substrate available for adaptation (Becker *et al*. 2011; Schmitz *et al*. 2011).

Acknowledgements

We thank the members of the Yi laboratory for discussion throughout the course of the research.

References

Anastasio AE, Platt A, Horton M *et al.* (2011) Source verification of mis-identified *Arabidopsis thaliana* accessions. *Plant Journal*, **67**, 554–566.

Becker C, Hagmann J, Muller J *et al.* (2011) Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*, **480**, 245–249.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, **153**, 51–68.

Calarco JP, Borges F, Donoghue MT *et al.* (2012) Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell*, **151**, 194–205.

Cortijo S, Wardenaar R, Colomé-Tatché M *et al.* (2014) Mapping the epigenetic basis of complex traits. *Science*, **343**, 1145–1148.

Des Marais DL, McKay JK, Richards JH, Richards JH, Sen S, Wayne T, Juenger TE (2012) Physiological Genomics of Response to Soil Drying in Diverse Arabidopsis Accessions. *The Plant Cell*, **24**, 893–914.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research*, **38**, W64–W70.

Dubin MJ, Zhang P, Meng D *et al.* (2015) DNA methylation in Arabidopsis has a genetic basis and shows evidence of local adaptation. *eLife*, **4**, e5025.

Elliott G, Hong C, Xing X *et al.* (2015) Intermediate DNA methylation is a conserved signature of genome regulation. *Nature Communications*, **6**, 6363.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–89.

Fraser HB (2013) Gene expression drives local adaptation in humans. *Genome Research*, **23**, 1089–1096.

Hagmann J, Becker C, Müller J *et al.* (2015) Century-scale methylome stability in a recently diverged *Arabidopsis thaliana* lineage. *PLoS Genetics*, **11**, e1004920.

Hannah MA, Wiese D, Freund S, et al. (2006) Natural Genetic Variation of Freezing Tolerance in Arabidopsis. *Plant Physiol*, **142**, 98–112.

Hancock AM, Brachi B, Faure N *et al.* (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*, **334**, 83–86.

Herman JJ, Spencer HG, Donohue K, Sultan SE (2014) How stable 'should' epigenetic modifications be? Insights from adaptive plasticity and bet hedging. *Evolution*, **68**, 632–643.

Heyn H, Moran S, Hernando-Herraez I *et al.* (2013) DNA methylation contributes to natural human variation. *Genome Research*, **23**, 1363–1372.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Horton MW, Hancock AM, Huang YS *et al.* (2012) Genome-wide patterns of genetic variation in worldwide *Arabidopsis thaliana* accessions from the RegMap panel. *Nature Genetics*, **44**, 212–216.

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 Research*, **24**, 1821–1829.

Johannes F, Porcher E, Teixeira FK *et al.* (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genetics*, **5**, e1000530.

Kalisz S, Purugganan MD (2004) Epialleles via DNA methylation: consequences for plant evolution. *Trends in Ecology & Evolution*, **19**, 309–314.

Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, **47**, 583–621.

Lamesch P, Berardini TZ, Li D *et al.* (2011) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research*, **40**, D1202–D1210.

Lasky JR, Des Marais DL, McKay JK, Richards JH, Juenger TE, Keitt TH (2012) Characterizing genomic variation of *Arabidopsis thaliana*: the roles of geography and climate. *Molecular Ecology*, **21**, 5512–5529.

Lasky JR, Des Marais DL, Lowry DB *et al.* (2014) Natural variation in abiotic stress responsive gene expression and local adaptation to climate in *Arabidopsis thaliana*. *Molecular Biology and Evolution*, **31**, 2283–2296.

Lasky JR, Upadhyaya HD, Ramu P *et al.* (2015) Genome-environment associations in sorghum landraces predict adaptive traits. *Science Advances*, **6**, e1400218.

Lister R, O'Malley RC, Tonti-Filippini J *et al.* (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, **133**, 523–536.

Liu QA (2013) The impact of climate change on plant epigenomes. *Trends in Genetics*, **29**, 503–505.

Liu L, Li Y, Tollefsbol TO (2008) Gene-environment interactions and epigenetic basis of human diseases. *Current Issues in Molecular Biology*, **10**, 25–36.

Long Q, Rabanal FA, Meng D *et al.* (2013) Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics*, **45**, 884–890.

Manel S, Joost S, Epperson BK *et al.* (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, **19**, 3760–3772.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, **18**, 1509–1517.

Moen EL, Zhang X, Mu W *et al.* (2013) Genome-wide variation of cytosine modifications between European and African populations and the implications for complex traits. *Genetics*, **194**, 987–996.

Nordborg M, Hu TT, Ishino Y *et al.* (2005) The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biology*, **3**, e196.

O'Connor TD, Fu W, Project NGES *et al.* (2015) Rare variation facilitates inferences of fine-scale population structure in humans. *Molecular Biology and Evolution* **32**, 653–660.

Oksanen J, Blanchet FG, Kindt R *et al.* (2015) vegan: Community Ecology Package. R package version 2.2-1.

Pavlidis P, Jensen JD, Stephan W, Stamatakis A (2012) A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Molecular Biology and Evolution*, **29**, 3237–3248.

Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Scheid OM (2010) Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *The Plant Cell*, **22**, 3118–3129.

Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, **87**, 2614–2625.

Platt A, Horton M, Huang YS *et al.* (2010) The scale of population structure in *Arabidopsis thaliana*. *PLoS Genetics*, **6**, e1000843.

Platt A, Gugger PF, Pellegrini M, Sork VL (2015) Genome-wide signature of local adaptation linked to variable CpG methylation in oak populations. *Molecular Ecology*, **24**, 3823–3830.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559–575.

Schmitz RJ, Schultz MD, Lewsey MG *et al.* (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, **334**, 369–373.

Schmitz RJ, Schultz MD, Urich MA *et al.* (2013) Patterns of population epigenomic diversity. *Nature*, **495**, 193–198.

Schubeler D (2015) Function and information content of DNA methylation. *Nature*, **517**, 321–326.

Segura V, Vilhjalmsson BJ, Platt A *et al.* (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics*, **44**, 825–830.

Sharbel TF, Haubold B, Mitchell-Olds T (2000) Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Molecular Ecology*, **9**, 2109–2118.

Shen X, De Jonge J, Forsberg SKG *et al.* (2014) Natural CMT2 variation is associated with genome-wide methylation changes and temperature seasonality. *PLoS Genetics*, **10**, e1004842.

Stroud H, Do T, Du J *et al.* (2014) Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. *Nature Structural & Molecular Biology*, **21**, 64–72.

Van Den Wollenberg AL (1977) Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika*, **42**, 207–219.

Walter H, Leith H (1960) *Klimadiagramm-Weltatlas*. Gustav-Fischer Verlag, Jena.

Zemach A, Kim MY, Hsieh PH *et al.* (2013) The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*, **153**, 193–205.

Zeng J, Konopka G, Hunt BG *et al.* (2012) Divergent whole-genome methylation maps of human and chimpanzee brains reveal epigenetic basis of human regulatory evolution. *American Journal of Human Genetics*, **91**, 455–465.

Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S (2006) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, **39**, 61–69.

Zomer RJ, Trabucco A, Bossio DA, Verchot LV (2008) Climate change mitigation: a spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems & Environment*, **126**, 67–80.

T.E.K. and S.V.Y. proposed the initial research question. T.E.K, J.R.L. and S.V.Y. designed research and analyzed the data. T.E.K, J.R.L., and S.V.Y. wrote and revised the manuscript.

Data accessibility

'Eurasian panel: Genomewide methylation maps for 140 *Arabidopsis thaliana* accessions from Schmitz *et al.* (2013) are archived at the NCBI GEO (Accession no. GEO43857), and variant calls for the same accessions were obtained from the NCBI SRA database (Accession no. SRA012474).

Swedish panel: whole-genome methylation maps for 148 accessions from across Sweden (Dubin *et al.* 2015) retrieved from GSE54292 and a SMP data set restricted to the 94 accessions with variants calls (Long *et al.* 2013) retrieved from http://plone.gmi.oeaw.ac.at/downloads/nordborg/data-release-for-massive-genomic-variation-and-strong-selection-in-arabidopsis-thaliana-lines-from-sweden.

Climate data for both the Eurasian and Swedish panel, as well as SMP and SNP data both before and after pruning, and C-DMR and CG-DMR data have been deposited in DRYAD (DOI: https://datadryad.org/resource/doi:10.5061/dryad.80442). In addition, we have supplied the distance matrices between accession sites, PCNM variables and RDA input files for R and their results.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1. Fraction of model variance explained by the first 30 RDA axes in different methylation contexts. The full model was calculated using each methylation context separately.

Table S1. Arabidopsis Ecotype accessions from Schmitz *et al.* 2013 and Dubin *et al.* 2015 used in climate association modeling.